Stem Cell Reports
Perspective
ISSCR
OPEN ACCESS

Investigating the feasibility and ethical implications of phenotypic screening using stem cell-derived tissue models to detect and manage disease

Alexander R. Harris,[1,*] Mary Jean Walker,[2] Frederic Gilbert,[3] and Patrick McGivern[4]
[1]Aikenhead Centre for Medical Discovery, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, VIC 3010, Australia
[2]Department of Politics, Media, and Philosophy, La Trobe University, Bundoora, VIC 3086, Australia
[3]School of Humanities, University of Tasmania, Hobart, TAS, Australia
[4]School of Humanities and Social Inquiry, University of Wollongong, Wollongong, NSW 2522, Australia
*Correspondence: alexrharris@gmail.com
https://doi.org/10.1016/j.stemcr.2022.04.002

Stem-cell-derived tissue models generated from sick people are being used to understand human development and disease, drug development, and drug screening. However, it is possible to detect disease phenotypes before a patient displays symptoms, allowing for their use as a disease screening tool. This raises numerous issues, some of which can be addressed using similar approaches from genetic screenings, while others are unique. One issue is the relationship between disease disposition, biomarker detection, and patient symptoms and how tissue models could be used to define disease. Other issues include decisions of when to screen, what diseases to screen for, and what treatment options should be offered.

Introduction

In order to understand complex diseases, scientists have typically used animal models displaying similar phenotypes (Wellbourne-Wood and Chatton, 2018; Harris et al., 2020). For example, in the case of epilepsy, acute seizure rodent models are used (Barker-Haliski et al., 2017). While these models have provided insights into biological mechanisms, differences between species limit their utility for understanding human disease and developing effective treatments. This deficit is even greater for neurological disorders, as animals do not display similar higher cognitive functions.

Stem cells allow for the creation of human-derived tissue models to investigate human development and disease and for drug development and screening (Amin and Paşca, 2018). Tissue models developed from individual patients enable the investigation of unique genetic factors affecting them. Models created from patients suffering epilepsy, schizophrenia, and Alzheimer's disease have replicated disease phenotypes such as epileptiform activity, altered neural structure, and formation of insoluble plaques, respectively (Chen et al., 2020; Falk et al., 2016; D'Souza et al., 2021).

Tissue models can be created from patients with or without known genetic abnormalities. For example, researchers recently differentiated stem cells derived from patients diagnosed with young-onset Parkinson's disease (YOPD) into dopaminergic neurons (Laperle et al., 2020). The tissue models displayed changes in protein expression compared with controls despite no known genetic mutations. It would now be possible to detect these changes in protein expression in models derived from people before they displayed clinical symptoms, demonstrating that tissue models can be used for phenotypic, as opposed to commonly used genetic, screening of disease. This method introduces great diagnostic promises but simultaneously raises scientific and ethical issues that must be addressed during its early research and development phase to ensure that research is performed appropriately and that research participants are treated fairly and can give sufficient informed consent. For instance, to establish the boundaries between normal and abnormal phenotypes within the model, we must ensure that models have been generated from individuals correctly classified with respect to disease. In the study just mentioned, disease models were obtained from patients diagnosed with YOPD. However, if the same models had been tested prior to patient diagnosis, models assigned to the control group would have displayed altered protein expression. Similarly, control individuals are not currently diagnosed with YOPD but may be in the future. Model validation therefore requires long-term follow up of trial participants. Incorrect group assignment can impact the study's statistical power, affecting publication and subsequent understanding and development of a model. Therefore, we require greater knowledge of what is an abnormal model phenotype, ensuring their correct classification as normal or diseased. There is also a chance that asymptomatic individuals in the control group will display the YOPD (or other disease) phenotype in their model. While this is good for diagnosing disease in people, it raises concerns around informing affected individuals of their likelihood of developing YOPD.

This perspective investigates the technical limitations and ethical and clinical implications of using stem-cell-derived tissue models for diagnosing diseases and evaluating treatment options (Figure 1) and details the potential and risks associated with this technology. As cell-programming techniques mature, using them on fetal or asymptomatic patient tissue for disease screening is becoming possible. This raises questions including how should disease be defined? What diseases should be screened for? Who would be suitable for testing? How would this screening be performed? And what types of interventions should be

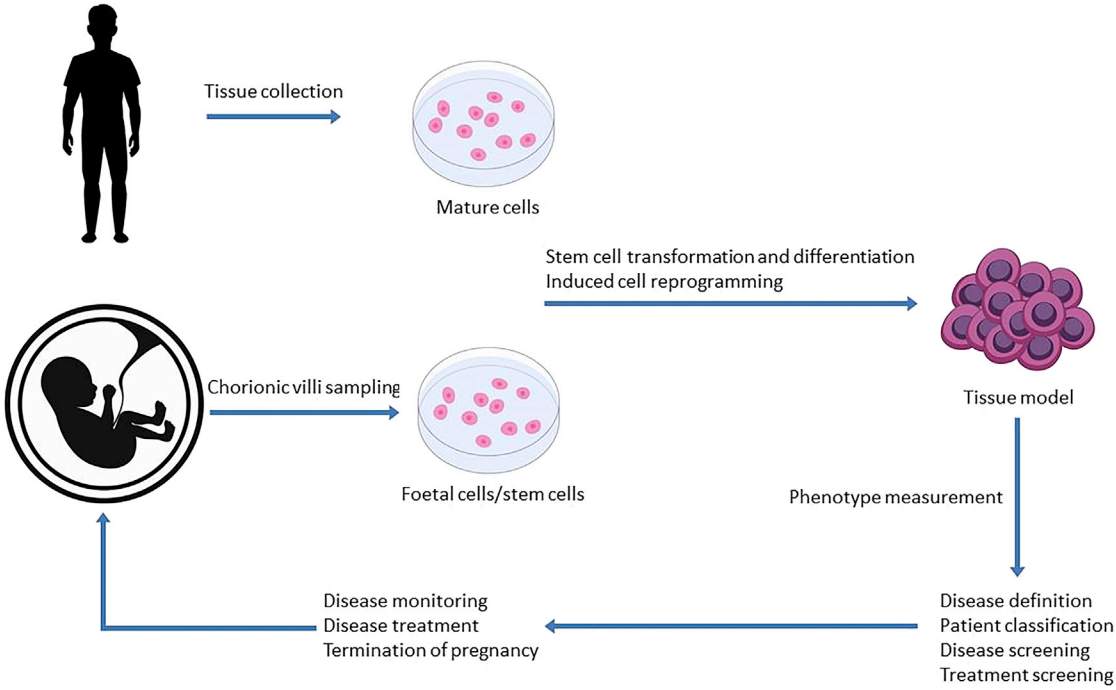

Figure 1. Potential method for disease screening using stem-cell-derived tissue models

provided? The ethical implications of phenotypic versus genetic screening and the clinical options available for undergoing screening and intervention are discussed. This perspective focuses on epilepsy as an example, as we have the most knowledge of this disease, but it does not imply that other diseases are not of equal importance.

Technical issues of phenotypic screening using stem cell-derived tissue models: Prenatal versus postnatal

Tissue models display certain phenotypes including protein and RNA expression, cellular and tissue structure, and electrophysiological activity. Ideally, these phenotypes replicate or correlate with normal, mature tissue and are not unnatural behaviors generated within the model. Abnormal model phenotypes may then be associated with a disorder. For instance, neural-tissue models from a schizophrenic sample may show unusual neural structure compared with a neurotypical sample, and an epileptic sample may display epileptiform activity not present in neurotypical tissue. These phenotypes may be visible directly in the model or may require a drug or electrical intervention to induce the abnormal phenotype.

Stem cell differentiation or direct induction to specific tissues (Zhang et al., 2013) could allow for the formation of tissue models from adults or fetuses before gestation week 20, the cut-off period for termination of a pregnancy in many jurisdictions. Models could also be created from cells taken from a preimplanted embryo used for *in vitro* fertilization (IVF). This would allow a preimplantation diagnosis with no risk to the mother or developing fetus, eliminating decisions on termination, but would only be available to women undergoing IVF.

While it is theoretically possible to form models from prenatal tissue, there are technical issues that require investigation. The ability to transform cells acquired from chorionic villi sampling (CVS) into a tissue model must be demonstrated. Given the risk to a pregnancy by performing CVS, it would be unethical to collect cells purely to investigate the efficacy of transforming tissue. However, if unused tissue was available following a prenatal diagnostic test, this would not increase the risk to a pregnancy, consistent with reuse of leftover human tissue when the alternative is it being discarded (Van Diest, 2002). A wider genetic population sample could be assessed by differentiating placental tissue obtained postpartum. Typical protocols around informed patient consent and handling of human tissue for research purposes would be required. Due to the high level of uncertainty in model behavior at this stage, early experimental work would only inform research protocols and not provide data for managing a pregnancy.

The relationship of model phenotypes to mature tissue must then be validated. While the disease phenotype may not be fully replicated in the model, certain behaviors may be correlated. For instance, a sample obtained from an epileptic patient may display abnormal electrophysiological activity but not the exact epileptiform waveform and location. This requires reproducibility of generating the

phenotypic behavior over sample repeats and across the disease population. The range of responses across neurotypical tissue must be quantified, and abnormal behavior associated with a disease must be significantly different (Bock et al., 2021). The tissue phenotypes that most strongly correlate with disease can then be quantified. Multiple phenotypes may increase detection accuracy, such as protein expression and electrophysiological behavior. The measurement reproducibility and significance as well as false-positive and -negative rates will then dictate the number of repeats required per sample, which will define the amount of tissue needed from CVS. These parameters may vary with disease and methodology, so appropriate validation and standardization must be undertaken.

CVS and amniocentesis are used for a final diagnostic genetic test for Down syndrome, as the accuracy is extremely high. The application of CVS for phenotypic screening may only be ethical if it has similar detection accuracy. Lower accuracy may make this screening procedure suitable only for very high-risk pregnancies or for adult screening. Further understanding of how phenotypic screening could be used requires knowledge of how genetic screening and diagnosis are currently used and their limitations.

Survey of key issues with currently accepted genetic screening and diagnosis of disease: Benefits and limitations

Genetic screening, prenatal diagnostic testing, and preimplantation genetic diagnoses are routinely undertaken prior or during pregnancy for the detection of genetic disorders such as Down syndrome, while adult screening is often performed for patient diagnosis or to determine parents' risk of passing on a disease to future offspring. Adult screening can be performed with blood or saliva samples, while prenatal diagnosis is more invasive. CVS takes tissue from the placenta, while amniocentesis samples take amniotic fluid (Perkins, 2017). These prenatal tests have a 1 in 16 chance of detecting an affected pregnancy, and 1 in 200 procedures can lead to miscarriage (Palomaki et al., 2011). A positive response from a prenatal diagnosis enables parents to prepare for the needs of an affected child, to make informed decisions regarding future family planning, or to make an informed choice about termination. With the availability of prenatal screening, the number of babies born with Down syndrome in Australia and the UK is decreasing (Diamandopoulos and Green, 2018).

The introduction of genetic screening and diagnosis has been generally favorable with the public (Henneman et al., 2013). Studies across multiple countries have shown high approval for screening of diseases (Aro et al., 1997; Vermeulen et al., 2014). Study participants believed people have a right to know if they or their future child are at risk of disease, and treatment costs may be reduced by early detection (Henneman et al., 2013). Concerns remained over data privacy potentially leading to discrimination and unapproved data usage, eugenics, and possible labeling of people with "good" and "bad" genes. It also raises issues for other family members for which the information may apply. As a result, participants felt testing should be widely available, particularly for family planning and during pregnancy, but should not be obligatory.

Despite procedure risk, prenatal screening and diagnosis have become common for many diseases and may increase with greater understanding of other genetic causes. However, many diseases are polygenic or have unknown causes not detectable by genetic screening, and the prediction of complex traits using a polygenic risk score (PRS) is very poor (Karavani et al., 2019). A major limitation of PRS approaches is that the datasets used to calculate risks are comprised predominantly of individuals of Northern European Caucasian descent, limiting their usefulness for other ethnic, demographic, and geographic populations (Adeyemo et al., 2021). These complex traits and diseases are detected and classified phenotypically, including protein expression, electrophysiological function, or patient symptoms. As a result, genetic screening is not suitable for detecting many diseases.

Epilepsy is one disorder that is diagnosed phenotypically, through abnormal brain activity and seizures (Fisher et al., 2014). Monogenic abnormalities account for 1%–2% of cases, with the vast majority being polygenic, caused by environmental factors or trauma. As a result, genetic screening for epilepsy will only detect a small number of cases; a significant number may be detected with a phenotypic screening method. While cases developing later in life from environmental factors, trauma, brain damage, and infection could not be screened for, around 75% of cases are caused by other factors and appear during childhood (Stafstrom and Carmant, 2015). These early-presenting cases strongly correlate with other comorbidities including neurological and developmental disorders, and sufferers have a high chance of having a close relative with epilepsy. Furthermore, ~30% of epileptic patients receive no benefit from currently available interventions (Kwan et al., 2010, 2011; Kwan and Brodie, 2000; Golyala and Kwan, 2017). These patients have a higher risk of death and injury, learning, concentration, and emotional issues, and stigma, which can impact child development and disease progression and management (Krauss and Sperling, 2011). Early diagnosis and treatment could have significant health, social, and learning benefits and prevent injuries or sudden unexpected death. This may greatly improve quality of life and reduce medical costs. In severe cases, terminating a pregnancy may be an option.

While the use of tissue models may enable detection of disease, it raises a new issue around how disease is defined.

Defining disease: Genetic versus phenotypic patient screening

Disease diagnosis of symptomatic patients is typically achieved by classifying various phenotypes including symptoms and diagnostic tests. This may lead to diagnosis of a specific disease or genetic abnormality. Genetic abnormalities can lead to a known disorder, like Down syndrome or monogenic epilepsy, being detected before symptoms arise. While there were concerns about genetic screening when this technology was first introduced, it is now generally accepted, and guidelines have been developed in response to issues raised (Henneman et al., 2013). Diagnosing asymptomatic people using phenotypic screening may not be as well accepted. One issue is the relationship between disease disposition, biomarker detection, and patient symptoms and how these define disease.

In most cases, diseases have no known genetic cause, and patients are diagnosed by clinical symptoms. However, there is a poor understanding of many normal and disease phenotypes. For instance, known genetic abnormalities can lead to epilepsy, which can be detected in asymptomatic people. But there is no quantifiable definition of an epileptic phenotype; epileptic brain activity is simply regarded as abnormal, and the patient suffers seizures (Fisher et al., 2014), but electrical activity varies between individuals and during brain development, and seizure measurement and prediction is poor. Epileptic patients can display abnormal brain activity without seizures, and self-reporting of seizures can differ significantly from electroencephalogram (EEG) recordings (Cook et al., 2013). Disease phenotypes can also arise and disappear over time; people may suffer seizures for a period of time during childhood or after certain events that then disappear for the remainder of their lives or reappear many years later (Fisher et al., 2014). This raises fundamental concerns over how disease is defined (Moynihan, 2011; Walker and Rogers, 2017a). Is someone "sick" when they are genetically predisposed to a disease, when they display certain phenotypes, or only when they begin suffering symptoms?

In many cases, there is no clear distinction between a healthy and sick person. Labeling someone suffering clinical symptoms as diseased without evidence of pathology or using a broad definition can improve their psychological well-being and access to support but may bias care giving and not result in treatment of symptoms (Bedson et al., 2004). There is an incentive to reduce disease thresholds to prevent people suffering unnecessary or permanent symptoms. However, pushing disease boundaries too wide can result in the majority of the population being diagnosed with a chronic condition. This can result in overtreatment of conditions that pose minimal risk, raising the possibility of harm from the treatment and unnecessary financial burdens. Overdiagnosis may be due to conflicts of interest in panels defining disease having links to drug companies, poor understanding of what early-detected phenotypes actually lead to harmful patient outcomes, physicians' fears of litigation for missed diagnoses, or increasingly sophisticated tests that identify subclinical pathologies. It has been argued that in light of such issues, there should be less patient screening (Reid, 2018), but this argument mainly relates to use of imaging to diagnose cancer and ignores the role confirmatory diagnostics such as biopsies can play prior to diagnosis and treatment; rather, the main issue is the disease definition.

While overdiagnosis is a risk, diagnosing a disease does not demand that treatment be administered, and, ultimately, more accurate disease definitions can only be achieved through greater understanding of their causes and trajectories.

Stem cell-derived models can provide more details about normal and abnormal phenotypes and a better understanding of the processes underlying diseases. This can lead to more precise patient diagnosis and, in the long run, better guidance on whether patients require treatment, what treatments should be administered, and when the best treatment stage is. It should also result in greater consistency in diagnoses and treatment outcomes between physicians.

As an example, the International League Against Epilepsy (ILAE) recently recommended 3 definitions of epilepsy (Fisher et al., 2014). A patient either suffers (1) at least two unprovoked or reflex seizures more than 24 h apart, (2) one unprovoked or reflex seizure with the probability of further seizures similar to the general recurrence risk (at least 60%) after two unprovoked seizures, occurring over the next 10 years, or (3) diagnosis of an epilepsy syndrome. The condition is considered resolved if the patient has suffered no further seizures after 10 years and is off antiseizure medications for 5 years. These definitions aim to align with how physicians and patients think about epilepsy and assist with treatment decisions. However, they raise issues around quantification of seizure risk and defining what an epilepsy syndrome is. Seizure risk is determined from population studies such as measurement of abnormal EEG studies after a first seizure, which subsequently correlates with a second seizure. However, an individual's risk can differ significantly from the overall population (Scott, 2003). Risk determined by EEG, magnetic resonance imaging (MRI), and blood tests can be subjective and be a poor predictor of disease development but may assist in diagnosing individual cases (Krumholz et al., 2007). An epilepsy syndrome is defined by various clinical features such as EEG, age of onset, and event duration, but the cause, disease trajectory, and treatment success vary

across people diagnosed with the same syndrome. Errors in self-reporting of clinical symptoms and the presence of comorbidities affect diagnosis accuracy (Wilden and Cohen-Gadol, 2012). It can also be difficult to diagnose neonates as the brain has not fully developed, affecting disease presentation, and the patient is unable to communicate (Stafstrom and Carmant, 2015). Stem cell-derived models may therefore provide more accurate individual-patient risks and redefine diseases around quantifiable phenotypes and drug responses rather than patient-reported symptoms. However, new issues will arise including redefining epilepsy in terms of tissue-model behavior rather than patient symptoms.

Patient screening may therefore be used to classify a disease, predict disease occurrence, or quantify a patient's risk of developing a disease. Regardless of disease trajectory, screening and early detection allows for the monitoring of disease progression, with the possibility of intervention and management to reduce risk of unexpected adverse events, and preventing or slowing disease progression. Stem cell-derived models may therefore play a role in disease definition.

Treatment choice following phenotypic screening using a stem cell-derived tissue model

If a disease phenotype is detected before patient symptoms arise, should treatment be administered? Currently, high blood pressure or cholesterol levels can be treated through medication or lifestyle changes. This reduces the risk of heart disease, heart attack, and stroke. In contrast, antiseizure medications are rarely prescribed unless someone has suffered seizures and been diagnosed with epilepsy (Fisher et al., 2014). Some neurologists administer treatment to low-risk patients after one seizure if the impact of a second seizure would be high (e.g., pilots or drivers who may lose their employment) (Wilden and Cohen-Gadol, 2012). This reflects the significant adverse effects associated with antiseizure medications (Liu et al., 2017). The prediction accuracy of the screening method is also relevant. Prenatal diagnostic testing of Down syndrome has a high prediction accuracy, while screening based on population risk factors for other diseases, including heart disease and many cases of epilepsy, are poor predictors of an individual developing the disease (Rockhill et al., 2000; Scott, 2003). This can be affected by reducing the level of risk for diagnosing a patient or by poorly defining groups in determining individual risk. A decision to treat should only be made where the screening prediction is sufficiently accurate.

If a tissue model indicated a risk of developing diseases such as epilepsy, drug or lifestyle interventions could be made before a seizure was detected. This could reduce the risk of injury, sudden unexpected death, or disease progression, particularly for patients unaware of their potential disorder. Unlike interventions for high blood pressure or cholesterol, those made from a tissue model may occur before any disease phenotypes are detected in the individual. The use of tissue models would allow drug or treatment screening to determine the most effective intervention in managing disease phenotypes before administration to the patient. Early intervention may result in improved outcomes and reduced side effects compared with later-stage intervention, which relies on diagnosing and monitoring patient symptoms and which may be complicated by comorbidities. Administering treatments for diseases before symptoms arise may therefore be consistent with reducing risk for susceptible patients.

There are also risks associated with early intervention. It may alter disease progression or symptoms and affect or prevent monitoring of disease progression. For instance, administering antiseizure medications may induce neural rearrangement, affecting seizure occurrence and prediction. Furthermore, similar phenotypes can arise from different diseases, so a specific treatment may be poorly targeted or have adverse outcomes. This is normally overcome by performing confirmatory tests or assessing patient response to various interventions, but this may not be possible if the patient is showing no disease phenotypes.

The decision to administer a treatment would be impacted by health-insurance coverage; however, insurance companies may consider covering preventative medication as a lower cost compared with later-stage interventions. Clinician fear of litigation may further encourage early intervention. As a result, overprescribing drug interventions before they provide clinical benefit or in place of potential lifestyle changes could occur.

Overall, the benefits of early intervention must be weighed against treatment costs and side effects. It may only be possible to estimate these at a population level and may be difficult to apply them to individuals. The development of new therapies targeting more specific disease cohorts with fewer side effects may shift the acceptable treatment stage and population.

Timing of phenotypic screening using a stem cell-derived tissue model: Prenatal versus postpartum

Diagnosing and treating disease phenotypes before symptoms are detected may lead to reduced risk and improved patient outcomes. Stem cell-derived models may allow prenatal or early-childhood phenotypic screening. However, there must be a balance of risk and benefit from a particular intervention for it to be clinically acceptable. Issues related to decisions about terminating pregnancies, early treatment, or new case-management methodology may depend on when a disorder is detected, raising questions about when someone should be screened.

One of the greatest concerns with prenatal screening, regardless of method, is the potential abuse of eugenics.

The current understanding of human development and the relationship between model phenotype and human characteristics is poor. So while phenotypic screening for highly abnormal conditions will be achievable in the short term, it is highly unlikely that detection of complex traits will be possible. However, rapid development of this field will increase discussion of this.

The main benefit of prenatal screening is the option of terminating an affected pregnancy. There are well-known concerns and procedures over the use of prenatal diagnostic testing of a fetus. The use of a tissue model for phenotypic screening will raise similar concerns but may extend them to previously undetectable conditions. In general, it may be acceptable to terminate a pregnancy if a tissue model predicted a child would develop severe mental and physical disabilities. In contrast, it would seem unacceptable to terminate a pregnancy if it predicted susceptibility to developing late-onset diseases such as Alzheimer's disease (Godard et al., 2003) (Condit, 2010). There will be a large gray area where some diseases may be considered neurodiverse (e.g., autism or attention-deficit hyperactivity disorder). Studies have shown that those with disabilities most people think reduce quality of life actually self-report similar quality of life to the nondisabled (Mackenzie and Scully, 2007). Labeling a fetus as disordered or susceptible to disease may unduly influence parents to terminate a pregnancy, perhaps ignoring factors such as poor detection accuracy, potential treatment options, or parents' limited ability to conceive. Therefore, the decision to screen, treat, or terminate a pregnancy is best made by the parents with the support of a counselor. Greater investigation of this issue will be required as the ability of phenotypic screening is understood.

It is possible that prenatal screening will provide no benefits compared with screening postnatally. Benefits of early detection must be balanced with the risk of affecting a pregnancy by performing CVS. Any clinically useful data must also be obtained before the typical cut-off period for termination of the pregnancy. Time spent screening can also increase parental anxiety. This may affect the pregnancy, so screening time and uncertainty should be minimized or avoided where possible. Genetic diagnosis of Down syndrome by CVS is only offered to women with increased risk of conceiving children with the syndrome (shown by previously having children with genetic abnormalities, family history of Down syndrome, advanced maternal age, or positive response from an earlier screening). Similar requirements could be set for prenatal phenotypic screening, limiting availability to women with an increased risk of conceiving children with a condition. A tissue model developed from an affected family member could provide a valuable control for assessing the fetal model. For a serious disorder, termination of the pregnancy may then be acceptable. In the absence of available treatments or options for termination, prenatal detection may still be desirable, as it allows parents to better prepare for an affected child or future pregnancy (Godard et al., 2003).

Other drug or lifestyle interventions may be administered to the mother during pregnancy to treat adverse phenotypes. These may alter fetal development, reducing the likelihood of developing the disease, its health impacts, or its progression. However, this can raise significant concerns. Our knowledge of fetal development is still relatively poor, and incorrect diagnoses and interventions can lead to significant harm to the fetus. There is a high risk that any intervention will have unintended side effects. The incidence of birth defects was 4%–7% when antiseizure medications were administered to pregnant women, nearly double the general-population rate (Liu et al., 2017). This kind of case has led to pharmaceutical companies avoiding the development of drugs for use during pregnancy and doctors advising women to avoid medications, limiting treatment options available during pregnancy and reducing the benefit of prenatal screening. The current risks of CVS and potential interventions, plus the limited treatment options available during pregnancy, make prenatal screening only suitable for the most serious diseases where termination is a viable option. As more treatment options are developed and the risks of intervention decrease, prenatal screening may become more acceptable for less-severe diseases.

Disease screening postnatally eliminates the risk of CVS, increases treatment options, and renders the time needed for screening irrelevant. Screening could be performed at any age. Early-childhood intervention may still affect brain development but would not result in birth defects. Interventions could be made at later life stages for late-onset diseases. Screening after birth would be performed to detect diseases before symptoms arise or for confirmation of a diagnosis at very early disease stages. Early treatment could then be instigated, preventing possible injuries or disease progression. However, a range of other ethical issues may arise during the clinical application of phenotypic screening.

Potential ethical issues in the clinical use of stem cell-derived tissue models: Drug versus disease screening

Stem cell-derived tissue models for drug screening are being developed for patients displaying disease symptoms (Chen et al., 2020). In these systems, a drug-testing battery would be administered to the model, and the drug displaying the best reduction of disease phenotypes would then be administered to the patient. For instance, a model displaying epileptiform activity may be created from a drug-resistant epileptic patient. A variety of antiepileptic drugs would be applied to the model to determine which was most effective at reducing the epileptiform activity. The

use of tissue models allows for testing a wider range of potential drugs in a shorter period than could be administered to the patient. This would allow for faster detection of the most effective treatment option, reducing risks of harm to the patient while unmedicated.

While the creation of stem cell-derived models for drug screening raises a range of general research ethics issues (ISSCR Guidelines for Stem Cell Science and Clinical Translation, 2021), new ethical concerns can arise during their clinical application (Walker et al., 2022). One concern is for patients already receiving some benefit from medication but the model predicts better disease management with another drug. In this case, there may be reluctance to change medication, particularly if the patient has spent large time periods unmedicated and suffered health or social impacts. Multiple drugs may display similar effects on the model, complicating treatment choice. Medically, there will need to be a shift from predicting disease management in a patient to disease management in a model. The models may also induce an overreliance on biomedical data for the clinician, reducing the impact of patient lifestyle and treatment side effects in choosing the most appropriate treatment option, and clinicians may fear litigation when choosing a treatment that contradicts the model's result. Disease screening with tissue models among asymptomatic people might raise further concerns. For instance, a routine screen might detect a susceptibility to epilepsy before the person has had a seizure. Under these circumstances, there must be a high level of trust in the screening model and appropriate guidance from medical practitioners to ensure the best patient outcome.

At the current state of development, it must be recognized these models are only part of a suite of diagnostic methods for guiding treatment and not an overriding decision-making tool. To ensure these screening tools are used correctly, clinicians must be provided with appropriate educational resources and training. Regulations and guidelines have been introduced to manage genetic screening (Godard et al., 2003). Similar regulations and guidelines will be required for phenotypic screening, including what to screen for, how it should be performed, and which patients should be screened and at what stage.

The screening process and subsequent diagnosis of disease or risk of disease may raise psychological harms for participants and result in counterproductive changes in behavior (Walker and Rogers, 2017b). For instance, the effects of an unexpected diagnosis of abdominal aortic aneurysm or prostate cancer through screening of asymptomatic patients were investigated (Cotter et al., 2017). While there were limited studies available for review, it was found that disease labeling of unexplained symptoms could reassure undiagnosed patients, but labeling asymptomatic patients could lead to psychological harms including shock, anxiety, fatalism, general distress, and burden about protecting others from worrying. In another screening study of hypertension, the screening process could induce hypertension, and patients who were newly diagnosed had increased absenteeism from work and depression even though they had no changes in symptoms (Pickering, 2006). False-positive diagnoses can also lead to long-term psychological harm (Brodersen and Siersma, 2013). Diagnosing a risk of developing disease can minimize the appreciation of environmental risks so that patients do not reduce other risk factors in disease development (Rockhill et al., 2000).

Screening could lead to unexpected diagnoses where screening for one disease detects another. This is a possibility with stem cell-derived models, where screening of protein expression may indicate susceptibility to a wide range of diseases (e.g. a patient being screened for YOPD may be predicted to be susceptible to epilepsy). There is potential for increased patient anxiety including around diseases with no treatment options. This can result in poor uptake of screening, as has been seen in genetic screening for Huntington's disease. It will be important to include a counselor in the screening process to assist individuals, as is already used when genetic screening.

Diagnosing a person with a disease can have social impacts including stigma and discrimination and create issues around personal relationships, insurance or welfare coverage, and employment opportunities (Dawson, 2011). It is critical that appropriate data integrity is applied to any screening process. The screening process will require a phenotype database for assessing patient data. This information could be included in genetic databases already available, such as GenBank. This will require significant investment in data management and agreement on what information should be collected. Any data must be de-identified before inclusion. Some of this information may be subject to intellectual property rights, affecting its availability for clinical use. For instance, a company may identify a particular phenotype as being a strong predictor of a disease or control the rights to a model required for measuring a phenotype.

Further questions surround who should be screened and at what disease stage. In the current early stage of stem cell model development, tissue is being created from symptomatic patients to understand disease etiology and trajectory, phenotype expression, and drug response. As models become validated for clinical use, they will form part of a battery of diagnostic tests. During this early-adoption phase, clinicians will need to be involved in the research to understand the models' results, accuracy, and limitations. Initial use of the models will most likely be for disease definition and patient classification, with limited impact on treatment choice except in repurposing drugs for

treatment-refractory patients. As the early use of these models will mainly test treatment-refractory patients, it will likely include patients with comorbidities who have severe intellectual disabilities and children with developmental disorders, raising issues around research among vulnerable populations including their capacity for informed consent (Nickel, 2006; Brazier and Lobjoit, 2005; Mackenzie et al., 2014). Later on, these models may be used in population-wide testing, requiring investigation of the relative cost and benefit of large-scale screening and resource limitations including lab facilities, trained lab and clinical staff, and counselors.

There are potential issues around equal access to stem cell-derived models. The costs associated with stem cell collection, isolation, conversion, and expansion, followed by development of analytical tests, is a time-consuming and expensive process that insurers may not cover. This can lead to a health disparity between those who can afford disease screening or not. Research funding and interest could focus on specific high-profile diseases, despite others potentially having greater impact on patients and society or being easier to detect and treat. This may push focus onto easily treatable diseases, reducing support for difficult cases. There may also be a focus from pharmaceutical companies on diseases that impact richer populations, have a higher incidence level in rich countries, or are considered chronic and require long-term drug use, thus generating greater profit.

The creation and use of stem cell-derived models requires highly trained and skilled scientists in a well-equipped facility. This will further limit the availability of these models to rich people and countries and result in a stratification of disease treatment and subsequent prevalence, as has already been observed in the rates of Down syndrome decreasing more in affluent countries (Diamandopoulos and Green, 2018). Varying acceptability of stem cell use across countries and in different religious or cultural groups may further affect the adoption of these types of models and subsequent disease prevalence.

Conclusion

Stem cells are providing new tools for understanding and treating disease. Our understanding of normal- and abnormal-tissue behavior is improving rapidly. This is leading to new techniques for creating specific tissues for modeling and drug screening. It is also possible to use these models to screen for disease. Tissue model phenotypes may correlate with patient phenotypes, allowing for their use to detect polygenic or idiopathic diseases. This may allow clinical interventions before a patient displays any symptoms or for prenatal screening. Many issues arising from phenotypic screening are similar to those occurring during genetic screening, and similar approaches should be applied. However, there is a risk of rushing into phenotypic screening too early. It appears that the risks of testing, the lack of understanding of phenotype relation to disease, and the effect of early treatment are currently very high. Therefore, further research is needed before phenotypic screening is made available. The continued development of new cell-reprogramming and tissue modeling methods are expected to further impact the speed, accuracy, risk, and efficacy of disease screening. The use of stem cell models for disease screening is therefore expected to become more likely, and, hence, their ethical use more critical.

AUTHOR CONTRIBUTIONS

The manuscript was conceptualized and drafted by A.R.H. The perspective was critically edited by M.J.W., F.G., and P.M.

CONFLICTS OF INTEREST

The authors declare no competing interests.

REFERENCES

Adeyemo, A., Balaconis, M.K., Darnes, D.R., Fatumo, S., Granados Moreno, P., Hodonsky, C.J., Inouye, M., Kanai, M., Kato, K., Knoppers, B.M., et al. (2021). Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. Nat. Med. *27*, 1876–1884.

Amin, N.D., and Pașca, S.P. (2018). Building models of brain disorders with three-dimensional organoids. Neuron *100*, 389–405.

Aro, A.R., Hakonen, A., Hietala, M., Lönnqvist, J., Niemelä, P., Peltonen, L., and Aula, P. (1997). Acceptance of genetic testing in a general population: age, education and gender differences. Patient Educ. Couns. *32*, 41–49.

Barker-Haliski, M.L., Johnson, K., Billingsley, P., Huff, J., Handy, L.J., Khaleel, R., Lu, Z., Mau, M.J., Pruess, T.H., Rueda, C., et al. (2017). Validation of a preclinical drug screening platform for pharmacoresistant epilepsy. Neurochem. Res. *42*, 1904–1918.

Bedson, J., McCarney, R., and Croft, P. (2004). Labelling chronic illness in primary care: a good or a bad thing? Br. J. Gen. Pract. *54*, 932–938.

Bock, C., Boutros, M., Camp, J.G., Clarke, L., Clevers, H., Knoblich, J.A., Liberali, P., Regev, A., Rios, A.C., Stegle, O., et al. (2021). The organoid cell atlas. Nat. Biotechnol. *39*, 13–17.

Brazier, M., and Lobjoit, M. (2005). Protecting the Vulnerable: Autonomy and Consent in Health Care (Routledge).

Brodersen, J., and Siersma, V.D. (2013). Long-term psychosocial consequences of false-positive screening mammography. Ann. Fam. Med. *11*, 106 LP–115.

Chen, Z., Rollo, B., Antonic-Baker, A., Anderson, A., Ma, Y., O'Brien, T.J., Ge, Z., Wang, X., and Kwan, P. (2020). New era of personalised epilepsy management. BMJ *371*, m3658.

Condit, C.M. (2010). Public attitudes and beliefs about genetics. Annu. Rev. Genomics Hum. Genet. *11*, 339–359.

Cook, M.J., O'Brien, T.J., Berkovic, S.F., Murphy, M., Morokoff, A., Fabinyi, G., D'Souza, W., Yerra, R., Archer, J., Litewka, L., et al.

(2013). Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. Lancet Neurol. *12*, 563–571.

Cotter, A.R., Vuong, K., Mustelin, L.L., Yang, Y., Rakhmankulova, M., Barclay, C.J., and Harris, R.P. (2017). Do psychological harms result from being labelled with an unexpected diagnosis of abdominal aortic aneurysm or prostate cancer through screening? A systematic review. BMJ Open *7*, e017565.

D'Souza, G.X., Rose, S.E., Knupp, A., Nicholson, D.A., Keene, C.D., and Young, J.E. (2021). The application of *in vitro*-derived human neurons in neurodegenerative disease modeling. J. Neurosci. Res. *99*, 124–140.

Dawson, A. (2011). Public Health Ethics: Key Concepts and Issues in Policy and Practice (Cambridge University Press).

Diamandopoulos, K., and Green, J. (2018). Down syndrome: an integrative review. J. Neonatal. Nurs. *24*, 235–241.

Van Diest, P.J. (2002). For and against: No consent should be needed for using leftover body material for scientific purposes. Br. Med. J. *325*, 648–651.

Falk, A., Heine, V.M., Harwood, A.J., Sullivan, P.F., Peitz, M., Brüstle, O., Shen, S., Sun, Y.-M., Glover, J.C., Posthuma, D., et al. (2016). Modeling psychiatric disorders: from genomic findings to cellular phenotypes. Mol. Psychiatry *21*, 1167–1179.

Fisher, R.S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J.H., Elger, C.E., Engel, J., Jr., Forsgren, L., French, J.A., Glynn, M., et al. (2014). ILAE Official Report: a practical clinical definition of epilepsy. Epilepsia *55*, 475–482.

Godard, B., ten Kate, L., Evers-Kiebooms, G., and Aymé, S. (2003). Population genetic screening programmes: principles, techniques, practices, and policies. Eur. J. Hum. Genet. *11*, S49–S87.

Golyala, A., and Kwan, P. (2017). Drug development for refractory epilepsy: the past 25 years and beyond. Seizure *44*, 147–156.

Harris, A.R., McGivern, P., and Ooi, L. (2020). Modeling emergent properties in the brain using tissue models to investigate neurodegenerative disease. Neuroscientist *26*, 224–230.

Henneman, L., Vermeulen, E., van El, C.G., Claassen, L., Timmermans, D.R.M., and Cornel, M.C. (2013). Public attitudes towards genetic testing revisited: comparing opinions between 2002 and 2010. Eur. J. Hum. Genet. *21*, 793–799.

Karavani, E., Zuk, O., Zeevi, D., Barzilai, N., Stefanis, N.C., Hatzimanolis, A., Smyrnis, N., Avramopoulos, D., Kruglyak, L., Atzmon, G., et al. (2019). Screening human embryos for polygenic traits has limited utility. Cell *179*, 1424–1435.e8.

Krauss, G.L., and Sperling, M.R. (2011). Treating patients with medically resistant epilepsy. Neurol. Clin. Pract. *1*, 14–23.

Krumholz, A., Wiebe, S., Gronseth, G., Shinnar, S., Levisohn, P., Ting, T., Hopp, J., Shafer, P., Morris, H., Seiden, L., et al. (2007). Practice Parameter: evaluating an apparent unprovoked first seizure in adults (an evidence-based review): [RETIRED]. Neurology *69*, 1996 LP–2007.

Kwan, P., and Brodie, M.J. (2000). Early identification of refractory epilepsy. N. Engl. J. Med. *342*, 314–319.

Kwan, P., Arzimanoglou, A., Berg, A.T., Brodie, M.J., Hauser, W.A., Mathern, G., Moshé, S.L., Perucca, E., Wiebe, S., and French, J. (2010). Definition of drug resistant epilepsy: consensus proposal by the ad hoc task force of the ILAE commission on therapeutic strategies. Epilepsia *51*, 1069–1077.

Kwan, P., Schachter, S.C., and Brodie, M.J. (2011). Drug-resistant epilepsy. N. Engl. J. Med. *365*, 919–926.

Laperle, A.H., Sances, S., Yucer, N., Dardov, V.J., Garcia, V.J., Ho, R., Fulton, A.N., Jones, M.R., Roxas, K.M., Avalos, P., et al. (2020). iPSC modeling of young-onset Parkinson's disease reveals a molecular signature of disease and novel therapeutic candidates. Nat. Med. *26*, 289–299.

Liu, G., Slater, N., and Perkins, A. (2017). Epilepsy: treatment options. Am. Fam. Physician *96*, 87–96.

Mackenzie, C., and Scully, J.L. (2007). Moral imagination, disability and embodiment. J. Appl. Philos. *24*, 335–351.

Mackenzie, C., Rogers, W.A., and Dodds, S. (2014). Vulnerability: New Essays in Ethics and Feminist Philosophy (Oxford University Press).

Moynihan, R. (2011). A new deal on disease definition. BMJ *342*, d2548.

Nickel, P.J. (2006). Vulnerable populations in research: the case of the seriously ill. Theor. Med. Bioeth. *27*, 245–264.

Palomaki, G.E., Kloza, E.M., Lambert-Messerlian, G.M., Haddow, J.E., Neveux, L.M., Ehrich, M., van den Boom, D., Bombard, A.T., Deciu, C., Grody, W.W., et al. (2011). DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet. Med. *13*, 913–920.

Perkins, A. (2017). The lowdown on Down syndrome. Nurs. Made Incred. Easy *15*, 40–46.

Pickering, T.G. (2006). Now we are sick: labeling and hypertension. J. Clin. Hypertens. *8*, 57–60.

Reid, L. (2018). Is an indistinct picture "exactly what we need"? Objectivity, accuracy, and harm in imaging for cancer. J. Eval. Clin. Pract. *24*, 1055–1064.

Rockhill, B., Kawachi, I., and Colditz, G.A. (2000). Individual risk prediction and population-wide disease prevention. Epidemiol. Rev. *22*, 176–180.

Scott, K.G. (2003). Commentary: individual risk prediction, individual risk, and population risk. J. Clin. Child Adolesc. Psychol. *32*, 243–245.

Stafstrom, C.E., and Carmant, L. (2015). Seizures and epilepsy: an overview for neuroscientists. Cold Spring Harb. Perspect. Med. *5*, a022426.

Vermeulen, E., Henneman, L., van El, C.G., and Cornel, M.C. (2014). Public attitudes towards preventive genomics and personal interest in genetic testing to prevent disease: a survey study. Eur. J. Public Health *24*, 768–775.

Walker, M.J., and Rogers, W. (2017a). Defining disease in the context of overdiagnosis. Med. Heal. Care Philos. *20*, 269–280.

Walker, M.J., and Rogers, W.A. (2017b). Diagnosis, narrative identity, and asymptomatic disease. Theor. Med. Bioeth. *38*, 307–321.

Walker, M.J., Nielsen, J., Goddard, E., Harris, A., and Hutchison, K. (2022). Induced pluripotent stem cell-based systems for personalising epilepsy treatment: research ethics challenges and new

insights for the ethics of personalised medicine. AJOB Neurosci. *13*, 120–131.

Wellbourne-Wood, J., and Chatton, J.-Y. (2018). From cultured rodent neurons to human brain tissue: model systems for pharmacological and translational neuroscience. ACS Chem. Neurosci. *9*, 1975–1985.

Wilden, J.A., and Cohen-Gadol, A.A. (2012). Evaluation of first nonfebrile seizures. Am. Fam. Physician *86*, 334–340.

ISSCR Guidelines for Stem Cell Science and Clinical Translation. www.isscr.org/policy/guidelines-for-stem-cell-research-and-clinical-translation.

Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., et al. (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. Neuron *78*, 785–798.