

Review

Understanding False Negative in Prenatal Testing

Mark I. Evans^{1,2,†}, Ming Chen^{3,4,5,6,*}  and David W. Britt^{1,†}

¹ Fetal Medicine Foundation of America, Icahn School of Medicine at Mt. Sinai, New York, NY 10029, USA; evans@compregen.com (M.I.E.); dwbrit01@louisville.edu (D.W.B.)

² Department of Obstetrics & Gynecology, Icahn School of Medicine at Mt. Sinai, New York, NY 10029, USA

³ Department of Genomic Medicine, Changhua Christian Hospital, Changhua 50046, Taiwan

⁴ Department of Obstetrics and Gynecology, National Taiwan University Hospital, College of Medicine, Taipei 10041, Taiwan

⁵ Department of Biomedical Science, Dayeh University, Changhua 51591, Taiwan

⁶ Department of Medical Sciences, National Tsing Hua University, Hsinchu 30013, Taiwan

* Correspondence: mingchenmd@gmail.com or 104060@cch.org.tw; Tel.: +886-4-7238595 (ext. 2323)

† All three authors contributed equally to this study.

Abstract: A false negative can happen in many kinds of medical tests, regardless of whether they are screening or diagnostic in nature. However, it inevitably poses serious concerns especially in a prenatal setting because its sequelae can mark the birth of an affected child beyond expectation. False negatives are not a new thing because of emerging new tests in the field of reproductive, especially prenatal, genetics but has occurred throughout the evolution of prenatal screening and diagnosis programs. In this paper we aim to discuss the basic differences between screening and diagnosis, the trade-offs and the choices, and also shed light on the crucial points clinicians need to know and be aware of so that a quality service can be provided in a coherent and sensible way to patients so that vital issues related to a false negative result can be appropriately comprehended by all parties.

Keywords: down syndrome screening; noninvasive prenatal testing; microarray; neural tube defect; mendelian screening panel



Citation: Evans, M.I.; Chen, M.; Britt, D.W. Understanding False Negative in Prenatal Testing. *Diagnostics* **2021**, *11*, 888. <https://doi.org/10.3390/diagnostics11050888>

Academic Editor: Ritsuko Kimata Pooh

Received: 28 April 2021

Accepted: 14 May 2021

Published: 17 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The COVID pandemic has publicized some of the difficulties in the interpretation of screening tests that clinicians and statisticians have been debating for decades. Anyone paying attention has had a crash course in understanding that a “test” is not always perfect, in fact, it does not always give the right answer, and that there are trade-offs in the accuracy and reliability of results [1]. Those trained in the discipline of screening will, of course, recognize the utilization of some of its basic principles to explain the current situation [2,3]. The majority of physicians and the general public seem confused by the apparent inability of the laboratories to give what was generally assumed to be simple yes/no answers to very important questions.

The same principles apply to all forms of screening. In prenatal genetics, common examples include aneuploidy, specifically, Down syndrome using multiple methodologies, Mendelian carrier screening, and pre-eclampsia [4–6]. On the gynecology side of obstetrics and gynecology, screening for malignancies of the cervix, uterus, ovary, and breast have developed rapidly (but are beyond the scope of this article) [7–9]. Before moving onto the nuances of emphasis that can be applied to how tests are designed and, equally-importantly, how cut-offs are arranged to favor either emphasis on sensitivity or emphasis on low false positive rates, we will first review some basics.

The first issue is to distinguish between screening and diagnostic tests. Simply put, screening tests only adjust odds, and diagnostic tests are meant to give a definitive answer (Table 1). Historically, there have been between seven and ten criteria that have been felt necessary to consider before deciding to screen for a condition [2,3,10]. We expand

upon these and recategorize them to include issues of infrastructure, training, availability, acceptability, cost and political interference that before the COVID epidemic might have been taken for granted by those establishing guidelines for screening tests (Table 2).

Table 1. Diagnostic vs. screening tests.

Screening Tests	Diagnostic Tests
Meant for everyone	Only done on “at risk” patients
Only adjust odds and do not give a definitive answer	Meant to give a definitive answer
Tests typically have little risk	Tests may have some procedural risks
Typically less expensive	Typically more expensive

Table 2. Criteria for effective screening and testing programs.

Disease	Screening	Testing	Intervention
High enough incidence	Good performance metrics	Good performance metrics	Beneficial intervention possible
	Availability and affordability of screening	Availability and affordability of testing	Availability and affordability of intervention(s) at different levels
	Acceptability of screening	Acceptability of testing	Acceptability of intervention(s) at different levels
Impairing or fatal	Capacity for follow-up and feedback	Capacity for follow-up and feedback	Benefits outweigh risks
	Adequate political support and coordination for public health	Adequate political support and coordination for testing	Adequate political support and coordination for interventions

Discussing these issues in any depth is beyond the scope of the present paper, but the context in which screening and testing takes place can no longer be taken for granted. In practice, not all tests being used, however, follow these guidelines, which can lead to disproportionate expectations (either positive or negative), expenditures, and complications from follow-up testing that are likely “unnecessary”. The goal of a screening program is to detect the maximum number of “affected” individuals for the least number screened false-positive. Specifically, a program is not judged by whether any particular patient’s problem can be identified [1–3]. Where to put the cut-off points in balancing these two is, by definition, arbitrary, but must be maintained to maximum efficiency.

Medicine, and particularly obstetrics, involves the repetitive use of both diagnostic and screening tests. Most patients and too many physicians do not understand the difference [1–3]. Diagnostic tests are meant to give a definitive answer, may have risks, may be expensive, and are only meant for patients at a risk high enough to warrant it. Conversely, screening tests are meant for “everyone”, and their aim is to divide a group with high-enough risk to warrant diagnostic testing from those who do not. They do not give definitive answers [1–3]. How well they do their job is defined by metrics of sensitivity, specificity, positive, and negative predictive value. These principles of evaluation were introduced into clinical practice in the 1970s by Galen and Gambino [1].

The performance characteristics establish the boundaries of a playing field and a scoring system within which tests compete for better ways to do things. Sensitivity and specificity as test properties are relatively well-known. Less often used are the criteria that summarize some of the critical trade-offs that clinicians face. One such trade-off is the relative number of true positive cases and false positive cases, which is reflected in the ratio of the two: True positives and false positives are defined as the Positive Likelihood Ratio (PLR). A second trade-off involves the relative number of false negatives and true negatives, reflected in the ratio of the two: False negatives and true negatives are defined

as the Negative Likelihood Ratio (NLR). Competitive approaches should have PLRs that are high and NLRs that are far less than one and need to be “minimally fractional”.

Receiver operating characteristic curves are typically used to compare the efficacy of tests or combinations of tests by directly examining the trade-off between True Positives (Sensitivity) and False Positives (1—Specificity). In situations where positive cases are very rare, so that the data are unbalanced, Precision-Recall Curves are more useful because they are not biased by the overwhelmingly large number of negative cases [11] and instead they focus more attention on the proportion of positive cases.

Clinicians are in the position of having to take all of this information and present it in an understandable way to explain not only the prior risk of the patient, but also how that risk is changing over time using Bayesian principles [12]. In our experience, clinicians are more likely to see value in a screening test that changes a priori risk from 1/500 to 1/20 than one that changes it from 1/100,000 to 1/500 even though the likelihood ratio of the former is improved 10 times more in the former than in the latter. An explanation for this would be that in the former, a change in clinical behavior might occur whereas such change is less likely with the latter. We also previously reported that patients “suddenly” at risk (e.g., abnormal aneuploidy screening) had a higher “state of anxiety” over the situation than those with the same level of risk, but it was because of non-emergent risks such as advanced maternal age [13].

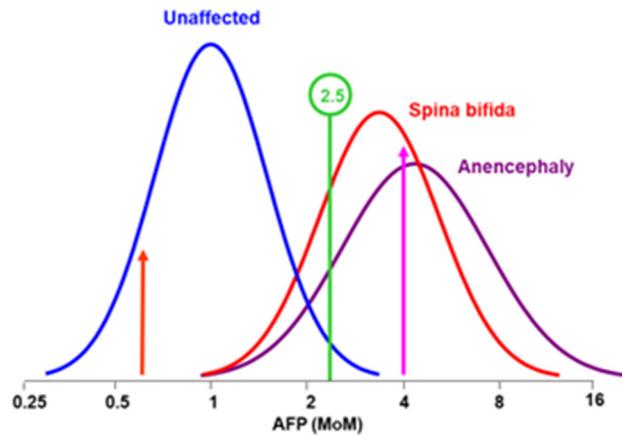
2. Choices

Given that, by definition, screening tests only adjust odds, there are options as to where to place the emphasis of the program. In our general approach to genetic counseling, we routinely tell patients that in most circumstances, for those in the “middle 98%”, it does not make a difference whether the patient has diagnostic testing for genetic abnormalities or not and everything will be fine. The question that we believe all patients need to understand is: “If you’re going to be wrong, which way would you rather be wrong? Would you rather take a small risk (e.g., a 2% risk of having a baby with a serious problem vs. a 0.1–0.12% risk of having a complication from a procedure in experienced hands) because you wanted to know that” [14–16]? We tell them that essentially the middle does not count.

The whole exercise then reduces to one question: “Tell me what you fear the most, then we can reduce that at the expense of the other options”. Unfortunately, encouraging patients to be active at the clinic level cannot be assumed to be representative across the board for at least three reasons. First, screening might be treated as part of routine medical care not “requiring” any patient decision-making processes [17], thus undermining the possibility that informed consent and shared decision-making would be part of the decision. Second, there may be a breakdown in shared decision-making owing to a various combination of class, racial-ethnic, or educational differences between genetic counselors and patients and the situation could create a collective fiction rather than shared decision-making [18]. Third, ideological differences may create at least a suspicion of genetic reasoning and analysis or outright hostility to it [19] or seek to create conditions in which genetic analysis assumes eugenic overtones [20,21].

The purpose of a screening program is not to maximize the correct diagnosis of any given patient. Rather it is to maximize the collective detection of problems for the least number of false positives. It can also be thought of as mimicking the ethical principle of proportionality, i.e., obtaining the most benefit for the least harm. A particularly good example was the use of maternal serum alpha-fetoprotein (MSAFP) for neural tube defects (NTDs). In the 1970s to the mid-1980s, the standard cut-off of 2.5 multiples of the median identified about 90% of NTDs for about a 5% false positive rate. Moving the cut-off to the right would increase the positive predictive value at the expense of sensitivity. Moving it to the left would increase the sensitivity but also create many more false positives (Figure 1). There is no absolute clarity as to where on the continuum the optimal cut-off point should be.

MATERNAL SERUM ALPHA FETOPROTEIN SCREENING: MOVING CUTOFFS FAR TO EITHER EXTREME



Courtesy: Dr. Howard Cuckle

Figure 1. Typical cut-off of 2.5 MOM identifies about 90% for a 5% false positive rate. Moving the cut-off to 4 MOM would significantly increase the positive predictive value in an abnormal, but would also result in many false negatives. Moving it far to the left (about 0.6 MOM) would increase the sensitivity by almost 100% but at the price of having a screen positive rate of nearly 70%.

Whether to maximize sensitivity or keep the screen positive number to a minimum involves not only mathematical analysis but clinical and ethical value judgements [2,3] (Figure 2). There are many components in the decision-making process for patients and couples as to whether they desire diagnostic testing.

CHOICES FOR CUT-OFF POINTS

FAVORING HIGH SPECIFICITY	FAVORING HIGH SENSITIVITY
Low incidence	High incidence
Low severity	High severity
Low positive predictive value	High positive predictive value
High cost of screening	Low cost of screening
High cost of testing	Low cost of testing
Hard/risky interventions	Easy/safe interventions
Low cost of disease care	High cost of disease care

Figure 2. Choices for cut-off points.

Given the ethical, religious and politically charged nature of prenatal diagnosis and its relationship to the abortion issue, it is not surprising that there are vast differences in the utilization of such technologies by regions of the world. Within large countries such as the United States, variations occur as a result of political affiliation, religious beliefs, social class and education, as well as multiple other personal and traditional “sociological” variables.

Policy makers have to make decisions about where to put cut-off points. Setting policies may be decided by the program directors, government officials, or insurance

companies. They may be uniform throughout a country or vary enormously even in the same environment. Alpha-fetoprotein (AFP) diagnosis from amniotic fluid was developed in 1972. At that time, the recurrence risk in the United Kingdom for a couple with one prior NTD was about 3–5% and, with 2 prior NTDs it was nearly 10%. Such patients were directly referred for amniocentesis even though its risks were thought to be nearly 2%, and because ultrasound was non-existent [22]. While recurrence risks were much higher than the primary incidence, 95% of all NTDs occurred in couples with no prior or family history [23,24]. The overall primary incidence of NTDs in Scotland was about 1/400. Given the high risk of procedures at the time, a screening approach was the only realistic public health measure to detect a highly significant or highly prevalent disorder. MSAFP was first introduced in 1973 as a screening test to determine who, among the supposedly low risk group, was at high risk [25].

The MSAFP cut-off was determined based upon conversations between the laboratory and the obstetrics department in Edinburgh to model what cut-off point was required to produce, by identifying and ranking the highest risk patients, the maximum number of “at risk” patients that could be accommodated by the clinical service at the hospital. Having the resources available is a relevant and rational contextual element is purely rational but assuming that the political process will operate in a way that is driven only by risk can no longer be taken for granted. Extending this argument, one can no longer assume that tests will not be reserved for those of higher status, or, on the other side, that tests will not be routine for those with lower status. At the most general level, neither micro-contexts nor macro-contexts [26] may be assumed to be immune from political influence.

Implementation has also been problematic. In 1988 we reported enormous variation in MSAFP screening in the Detroit area such that the same specimen could be interpreted by laboratories across the full spectrum from being very high risk to very low risk. Likewise, some laboratories stated the need for certain parameters to be known (such as maternal weight and ethnicity), but then did not insist that they be provided, or even include them in their analysis [27]. It took decades to coalesce on quality control measures to ensure better standardization of the accuracy of measurements and the relationship between the raw measurement and its clinical interpretation. In the 1970s, it was common belief that laboratories in different cities needed to have their own reference ranges for MSAFP because the populations were “different” [28]. Eventually, it was determined that the discordance was, to an exceptionally large extent, a function of different laboratory methodologies. When standardized across the country the differences between cities completely disappeared. Racial and ethnic differences still remained, but they also did not vary across cities.

In the mid-1980s, as MSAFP screening moved from just NTDs (high MSAFP) to adding Down syndrome and Trisomy 18 (low MSAFP), assay changes were required to increase the accuracy of low-level measurements [29]. One effect was a distribution shift such that the percentage of cases with high MSAFPs fell from about 5% to 2.5%. Given that the majority of NTD cases had MSAFP values over 4.0 multiple of the median (MoM), the change resulted in very few additional false negatives and simultaneously improved the positive predictive value of those cases at risk [30]. Increasing the quality of ultrasound technology and clinical expertise also mitigated the risks of missed MSAFP screens [31].

Likewise, as MSAFP expanded to double (MSAFP + total human chorionic gonadotropin (hCG), then with some centers substituting free β hCG for total), triple (adding estriol), and eventually quad (adding inhibin), there were many arguments over the best combination with most arguing about estriol [32]. Arguments between the “double” vs. the “triple” supporters reached almost religious fervor. (We belonged to the “double” religion). Ultimately, analysis showed that whereas variables such MSAFP and the hCGs had stable coefficients of variation (COV) at around 5–7%, the estriol assay was much less accurate [32]. Some laboratories reported a similar 5–7% COV and demonstrated that estriol added value. However, for many laboratories, the COV was as much as 30%, rendering estriol useless or worse [32].

With the introduction of nuchal translucency (NT) measurements, the COV problem became considerably worse. In order to control for highly variable quality and accuracy, the Fetal Medicine Foundation in London developed training and screening reviews and certifications and showed increasing quality world-wide [33]. Later the Nuchal Translucency Quality Review (NTQR) program of the Society for Maternal Fetal Medicine in America introduced training programs in the United States. We reported that the Fetal Medicine Foundation program seemed to generate higher quality control than NTQR which led to higher sensitivities and specificities [34,35]. We developed an approach to “handicap” less proficient providers by modulating how much those measurements counted in the algorithm [36]. It worked but never made it into the mainstream as the concurrent development of cell free fetal DNA came into practice whose protocols almost eliminated the use of NT measurements for Down syndrome screening in the practice landscape of prenatal diagnosis and obstetrics.

The two situations above (COV of estriol and NT quality) illustrate the conundrum of trying to improve the statistical performance of screening by insisting upon quality controls. It can be hard in the laboratory; it is much harder for more subjective variables, e.g., ultrasound NT measurements, with multiple clinicians involved [33–36]. Each provider has their own strengths and weaknesses, and there is well known variability among providers [37]. These types of situations emphasize further issues of gradient between absolute conformity of all care and widespread individuality that can elevate some providers but reveal weaknesses at the same time. The former elevates weaker providers but inhibits excellence at the top end.

Once Down syndrome became the primary focus of screening, multiple generations of improvements have occurred. The shifts from MSAFP to double, triple, and quadruple screening in the second trimester and NT, free β hCG, pregnancy associated plasma protein A (PAPP-A), combined screening, cell free fetal DNA (cffDNA) in various forms in the first trimester have all had the underpinnings of simultaneously improving both sensitivity and specificity [38–40]. Ironically, however, the improved screening for Down syndrome has come at the cost of increased failures to detect other problems that could have been found by using high quality ultrasounds and diagnostic procedures such as chorionic villus sampling with improved laboratory tests such as microarrays that can find, particularly in younger patients, 10 times the abnormalities that can be done by cffDNA [14–16].

Changes in the perceived efficacy of tests can have wide ranging effects upon the practice of medicine and society. On the positive side, improvements in Mendelian Screening have allowed the identification of at-risk couples to occur BEFORE they have experienced the impact of an affected child. Likewise, most fetal cardiac anomalies can now be detected prenatally allowing births to be moved to centers capable of advanced care rather than having to have panic neonatal transfers or no transfers at all [41,42]. Unfortunately, the rate of incorporation of new approaches is highly variable among industries. Medicine has historically been on the slower end of the spectrum and obstetrics on the slower end of medicine [43]. NT screening is a good example. It was adopted much faster in the UK and Europe than in the USA [37,43]. Likewise, there is a pattern in which there is first the development of new technologies followed by diffusion out to the community. As new methods expand, the utilization increases, but complications can increase rapidly until there is time for community understanding and education [44].

The sociological and societal implications of such understanding have dramatic implications for the practice of medicine and the health of the population. Due to the increased “screening” for Down syndrome there has been diminished utilization of procedures that we have described as an epidemic of MISSED abnormalities because of cffDNA. The public health effect has meant more unintended babies with serious disorders that, alongside other issues, have expensive care requirements [14–16].

Even getting past technical, financial, implementation issues, and quality controls, there are still philosophical issues that drive the determination of a cut-off point. The more serious the disorder, the more likely that any typical patient’s “tolerance” for missing a

case would be very low [45,46]. This would create an impetus to put the cut-off point “to the left” on the distribution curve which would maximize sensitivity at the expense of increased false positives. Likewise, with disorders with a very low incidence or those in which morbidity and mortality might be moderate, it would be reasonable to move the cut-off to the right to minimize the number of patients undergoing follow-up diagnostic procedures [1–3] (Figure 2).

We have previously modeled the detection of genetic disorders comparing those identifiable by cffDNA versus diagnostic testing using chorionic villus sampling or amniocentesis with enhanced genetic testing using array comparative genomic hybridization (aCGH or microarrays) [14–16]. We reported that the yield of detected abnormalities was about 10 times greater using the diagnostic testing vs. screening. While this is to some degree an apples vs. oranges comparison (screening vs. diagnostic), it does highlight the trade-off of simplicity and lower cost versus comprehensiveness and greater expense for the detection and management of rare but significant outcomes. More direct examples within the screening world would be karyotype vs. microarray, small ethnically derived Mendelian screening panels vs. larger pan-ethnic panels, and Pap smears and mammograms vs. BRCA and Lynch syndrome molecular panels [7–9].

Some prenatal diagnosis centers are known for their procedure acumen. Patients who want “definitive” answers tend to go to such centers, often directed by referring physicians sensitive to their patient’s desires [14]. Conversely, other centers are known for a limited use of diagnostic procedures. To a degree, everyone “votes with their feet” as to where they go. A continuing challenge is assuring that patients have sufficient access and understanding to make informed “foot” (certainly autonomous) choices.

For some parameters, such as advanced maternal age, there are at least de facto standards as to what the borderline is. In the United States, advanced maternal age was 40 in the early 1970s, then fell to 38, and has been 35 for over three decades [3]. However, screening has used the cut-off as a starting point. For methods such as MSAFP, multiple markers and ultrasounds in the second trimester, and combined screening and nuchal translucencies in the first trimester, statistically use a “likelihood ratio” to multiply (\times) the screening result with the a priori risk of achieving a risk which is then compared to the a priori risk [34,36]. The situation with cffDNA is more complicated as some laboratories use an absolute, (i.e., we see it, or we do not), and others base their risk prediction estimate on the background risk as defined by maternal age. Some vary their risk certainty by the fetal fraction of the specimens, and others do not [47]. These variations in style occur as physicians and clinics adapt to the needs, preferences and cultures of the populations they serve. The systemic challenge is to continually update both training and the diffusion of more effective technologies.

cffDNA testing was originally publicized by its developers to be a diagnostic test. The acronym NIPD for noninvasive prenatal diagnosis was commonly used. However, it soon became clear that “NIPD” was only a screening test based upon calculation no matter which methodology was chosen [48].

Fetal cell testing was stopped when cffDNA was developed. Recently, some researchers have tried to revive the cell-based non-invasive prenatal testing (cbNIPT). There are theoretical advantages of cells vs. cffDNA. Cell-based methods would give a more direct confirmation of fetal vs. maternal DNA and could have faster whole-genome applications. If so, the cell-based testing could be the solution to creating a diagnostic rather than a screening test and thus be capable of reducing false negatives inherent in cffDNA prenatal testing.

Some early cbNIPT studies (1970s and 1980s) were based upon capturing trophoblasts, which intuitively cannot solve the limitation of fetoplacental mosaicism [49,50]. In the 1990s and early 2000s, most approaches, including the National Institute of Child Health’s “NIFTY study” used nucleated erythrocytes, but the technologies were not robust enough [51]. Newer methods using fetal nucleated red blood cells (fnRBCs) might be able to solve the limitation of fetoplacental mosaicism which is inherent in both trophoblast-based cbNIPT

and cffDNA testing. However, the possible revival of cbNIPT, even with the utility of fnRBCs instead of trophoblasts as the diagnostic target, will be built upon the prerequisite to solve the long-existing limitations of the technologies themselves: questionable reproducibility, consistency, scalability, and even reliability should be improved first. Automation by incorporating artificial intelligence (AI) to reduce manual work may be a way to help, and the progress in the highly similar field of circulating tumor cells (CTCs) can be referenced [52]. Much more work, including large-scale validation studies, is needed to verify its feasibility [53,54]. Supposing such efforts by capturing fnRBCs materialize, the problem of true fetal mosaicism is still beyond the scope of such technology.

We raise this issue to highlight the next generation “tug of war” between screening versus diagnostic approaches that we have previously addressed between cffDNA and microarrays [14–16]. Well-informed genetic counseling will be critical before utilizing any emerging new genetic testing, either screening or diagnostic. Considerable evidence suggests that the vast majority of obstetricians world-wide are not well-versed in the new technologies [14–16,47,55]. The more reliable a test is, the less dependent health care delivery will be on resources that do not exist for a very high-level clinical interpretation of data.

3. Conclusions

A diagnostic test is meant to give a definitive answer. Screening tests only alter odds. There are many assumptions and decisions that have to be made arbitrarily by the program directors that profoundly influence the outputs and can be swayed towards one extreme (maximizing sensitivity) or the other (minimizing screen positives). Even minor shifts in screening program philosophy and practice can have enormous public health implications.

Most programs do not publicly report these underlying decisions, or how their operations conform or vary from other laboratories in their area. Most programs likewise do not declare their endpoints in situations in which outcomes are not clear-cut and how such end point uncertainty colors their predictive values. There is no likelihood that there will be any uniformly agreed-upon strategy for such decisions, but clinicians should at least be aware of these issues and how they affect the performance of all screening tests. Understanding the scientific and sociological contexts of false negative results, in addition to false positive results, will be very important when clinicians are interpreting the results of prenatal screening and diagnostic tests and thereby can offer counseling to their patients with acceptable and decent quality. The misconception of confusing a screening test with a diagnostic one will bring negative impacts to the overall obstetric practice and the populations they serve, as we have seen in the recent history when cffDNA-based NIPT was introduced since 2011 [14–16].

Author Contributions: M.I.E., M.C. and D.W.B.; writing—original draft preparation, M.I.E., M.C. and D.W.B.; writing—review and editing, M.I.E., M.C. and D.W.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Changhua Christian Hospital, grant number 109-CCH-PRJ-117.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Galen, R.S.; Gambino, S.R. *Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses*; Wiley: New York, NY, USA, 1975.
2. Evans, I.M.; Krivchenia, X.; Wapner, R.; Depp, R. Principles of Screening. *Clin. Obstet. Gynecol.* **2002**, *45*, 657–660. [[CrossRef](#)]
3. Evans, M.I.; Eden, R.D.; Britt, D.W.; Evans, S.M.; Schifrin, B.S. Re-conceptualizing fetal monitoring. *Eur. J. Gynecol. Obstet.* **2019**, *1*, 10–17.
4. Haque, I.S.; Lizarin, G.A.; Kang, H.P.; Evans, E.A.; Goldberg, J.D.; Wapner, R.J. Modeled fetal risk of genetic diseases identified by expanded carrier screening. *JAMA* **2016**, *316*, 734–742. [[CrossRef](#)] [[PubMed](#)]
5. Rolnik, D.L.; Wright, D.; Poon, L.C.; O’Gorman, N.; Syngelaki, A.; de Paco Matallana, C.; Akolekar, R.; Cicero, S.; Janga, D.; Singh, M.; et al. Aspirin versus placebo in pregnancies at high risk for preterm preeclampsia. *NEJM* **2017**, *377*, 613–622. [[CrossRef](#)]

6. Hashiloni-Dolev, Y.; Nov-Klaiman, T.; Rza, A. Pandora's pregnancy: NIPT, CMA, and genetic sequencing—A new ear for prenatal genetic testing. *Prenat. Diagn.* **2019**, *39*, 859–865. [[CrossRef](#)]
7. Committee on Practice Bulletins-Gynecology. Practice Bulletin number 179: Breast Cancer risk assessment and screening in average risk women. *Obstet. Gynecol.* **2017**, *130*, e1–e16. [[CrossRef](#)] [[PubMed](#)]
8. Pellat, A.; Netter, J.; Perkins, G.; Cohen, R.; Coulet, F.; Parc, Y.; Svrcek, M.; Duval, A.; André, T. Lynch syndrome: What is new? *Bull. Cancer* **2019**, *106*, 647–655. [[CrossRef](#)] [[PubMed](#)]
9. Staples, J.N.; Duska, L.R. Cancer screening and prevention highlights in gynecologic cancer. *Obstet. Gynecol. Clin. N. Am.* **2019**, *46*, 19–36. [[CrossRef](#)]
10. Wilson, J.M.G.; Jungner, G. *Principles and Practice of Screening for Disease*; World Health Organization: Geneva, Switzerland, 1968.
11. Saito, T.; Rehmsmeier, M. The precision recall plot is more informative than the ROC plot when evaluating classifiers on imbalanced datasets. *PLoS ONE* **2015**, *10*, e0118432. [[CrossRef](#)] [[PubMed](#)]
12. Ogino, S.; Wilson, R.B. Bayesian analysis and risk assessment in genetic counseling and testing. *J. Med. Diagn.* **2004**, *6*, 1–9. [[CrossRef](#)]
13. Evans, M.I.; Bottoms, S.F.; Carducci, T.; Grant, J.; Belsky, R.L.; Solyom, A.E.; Quigg, M.H.; LaFerla, J.J. Determinants of altered anxiety following abnormal maternal serum alpha-fetoprotein screening. *Am. J. Obstet. Gynecol.* **1988**, *159*, 1501–1504. [[CrossRef](#)]
14. Evans, M.I.; Wapner, R.J.; Berkowitz, R.L. Noninvasive prenatal screening or advanced diagnostic testing: Caveat emptor. *Am. J. Obstet. Gynecol.* **2016**, *215*, 298–305. [[CrossRef](#)] [[PubMed](#)]
15. Evans, M.I.; Andriole, S.; Curtis, J.; Evans, S.M.; Kessler, A.A.; Rubenstein, A.F. The epidemic of abnormal copy number variant cases missed because of reliance upon noninvasive prenatal screening. *Prenat. Diagn.* **2018**, *38*, 730–734. [[CrossRef](#)] [[PubMed](#)]
16. Evans, M.I.; Evans, S.M.; Bennett, T.A.; Wapner, R.J. The price of abandoning diagnostic testing for cell-free fetal DNA screening. *Prenat. Diagn.* **2018**, *38*, 243–245. [[CrossRef](#)] [[PubMed](#)]
17. Santalahti, P.; Hemminki, E.; Aro, A.R.; Helenius, H.; Ryyänen, M. Participation in prenatal screening tests and intentions concerning selective termination in Finnish medical care. *Fetal. Diagn. Ther.* **1999**, *14*, 71–79. [[CrossRef](#)]
18. Press, N.; Browner, C.H. Characteristics of women who refuse an offer of prenatal diagnosis: Data from the California maternal serum alpha fetoprotein blood test experience. *Am. J. Med. Genet.* **1998**, *78*, 433–445. [[CrossRef](#)]
19. Matthew, D.B. Two threats to precision medical equity. *Ethn. Dis.* **2019**, *29*, 629–640. [[CrossRef](#)]
20. Fletcher, J. The long view: How genetic discoveries will aid health care reform. *J. Women's Health* **1998**, *7*, 817–823. [[CrossRef](#)]
21. Duster, T. *Backdoor to Eugenics*, 2nd ed.; Routledge: New York, NY, USA, 2003.
22. Wald, N.; Cuckle, H. Amniotic fluid alpha-fetoprotein measurement in antenatal diagnosis of anencephaly and open spina bifida in early pregnancy. *Lancet* **1979**, *1*, 685–688. [[CrossRef](#)]
23. Gastel, B.; Haddow, J.E.; Fletcher, J.C.; Neale, A. *Maternal Serum Alpha-Fetoprotein: Issues in the Prenatal Screening and Diagnosis of Neural Tube Defects*; U.S. Government Printing Office: Washington, DC, USA, 1980.
24. Lupo, P.J.; Agopian, A.J.; Castillo, H.; Castillo, J.; Clayton, G.H.; Dosa, N.P.; Hopson, B.; Joseph, D.B.; Rocque, B.G.; Walker, W.O.; et al. Genetic epidemiology of neural tube defects. *J. Pediatr. Rehabil. Med.* **2017**, *10*, 189–194. [[CrossRef](#)]
25. Brock, D.J.H.; Bolton, A.E.; Monaghan, J.M. Prenatal diagnosis of anencephaly through maternal serum alpha-fetoprotein measurements. *Lancet* **1973**, *2*, 923–924. [[CrossRef](#)]
26. Bronfenbrenner, U. Contexts of child rearing: Problems and prospects. *Am. Psychol.* **1979**, *34*, 844–850. [[CrossRef](#)]
27. Evans, M.I.; Belsky, R.L.; Greb, A.E.; Dvorin, E.; Drugan, A. Wide variation in maternal serum alpha-fetoprotein (MSAFP) reports in one metropolitan area: Concerns for the quality of prenatal testing. *Obstet. Gynecol.* **1988**, *72*, 342–345.
28. Braoudakis, G.; Marguerat, P.; Marazzi, A.; Gutzwiller, F. Interlaboratory variation in the prenatal screening for neural tube closing defects. *Soz. Präventivmed.* **1984**, *29*, 207–208. [[CrossRef](#)]
29. Macri, J.N.; Kasturi, R.V.; Krantz, D.A.; Hu, M.G. Maternal serum alpha-fetoprotein screening: II. Pitfalls in low volume decentralized laboratory performance. *Am. J. Obstet. Gynecol.* **1987**, *156*, 533–535. [[CrossRef](#)]
30. Reichler, A.; Hume, R.F., Jr.; Drugan, A.; Bardicef, M.; Isada, N.B.; Johnson, M.P.; Evans, M.I. Risk of anomalies as a function of level of elevated maternal serum alpha-fetoprotein. *Am. J. Obstet. Gynecol.* **1994**, *171*, 1052–1055. [[CrossRef](#)]
31. Schell, D.L.; Drugan, A.; Brindley, B.A.; Zador, I.E.; Johnson, M.P.; Schwartz, D.B.; Evans, M.I. Combined ultrasonography and amniocentesis for pregnant women with elevated maternal serum alpha-fetoprotein: Revising the risk estimate. *J. Reprod. Med.* **1990**, *35*, 543–546.
32. Evans, M.I.; Hallahan, T.W.; Krantz, D.; Galen, R.S. Meta-Analysis of first trimester Down Syndrome screening studies: Free beta hCG significantly outperforms intact hCG in a multi-marker protocol. *Am. J. Obstet. Gynecol.* **2007**, *196*, 198–205. [[CrossRef](#)]
33. Nicolaides, K.H. Turning the pyramid of prenatal care. *Fetal. Diagn. Ther.* **2011**, *29*, 183–196. [[CrossRef](#)] [[PubMed](#)]
34. Evans, M.I.; Krantz, D.A.; Hallahan, T.W.; Sherwin, J.E. Undermeasurement of nuchal translucencies: Implications for screening. *Obstet. Gynecol.* **2010**, *116*, 815–818. [[CrossRef](#)]
35. Evans, M.I.; Krantz, D.; Hallahan, T.; Sherwin, J. Impact of nuchal translucency credentialing by the FMF, the NTQR or both on screening distributions and performance. *Ultrasound Obstet. Gynecol.* **2012**, *39*, 181–184. [[CrossRef](#)] [[PubMed](#)]
36. Evans, M.I.; Cuckle, H.S. Performance Adjusted Risks: A method to improve the quality of algorithm performance while allowing all to play. *Prenat. Diagn.* **2011**, *31*, 797–801. [[CrossRef](#)]
37. Evans, M.I. Overcoming Militant Mediocrity. *Am. J. Obstet. Gynecol.* **2008**, *198*, 656–661. [[CrossRef](#)]

38. Evans, M.I.; Sonek, J.D.; Hallahan, T.W.; Krantz, D.A. Cell-free fetal DNA screening in the USA: A cost analysis of screening strategies. *Ultrasound Obstet. Gynecol.* **2015**, *45*, 74–83. [[CrossRef](#)] [[PubMed](#)]
39. Van der Meij, K.R.M.; Sistermans, E.A.; Macville, M.V.E.; Stevens, S.J.C.; Bax, C.J.; Bekker, M.N.; Bilardo, C.M.; Boon, E.M.J.; Boter, M.; Diderich, K.E.M.; et al. TRIDENT 2: National implementation of genome wide non invasive prenatal testing as a first tier screening test in the Netherlands. *Am. J. Hum. Genet.* **2019**, *105*, 1091–1101. [[CrossRef](#)]
40. Bunnik, E.M.; Kuipers, A.K.; Galjaard, R.J.H.; de Beaufort, I.D. Should pregnant women be charged for non-invasive prenatal screening? Implications for reproductive autonomy and equal access. *J. Med. Ethics* **2020**, *46*, 194–198. [[CrossRef](#)] [[PubMed](#)]
41. Britt, D.W. Psychosocial issues in Prenatal Diagnosis. In *Prenatal Diagnosis*; Evans, M.I., Johnson, M.P., Yaron, Y., Drugan, A., Eds.; McGraw Hill Professional: New York, NY, USA, 2006; pp. 701–706.
42. Donofrio, M.T. Predicting the future: Delivery room planning of congenital heart disease diagnosed by fetal echocardiography. *Am. J. Perinatol.* **2018**, *35*, 549–552. [[CrossRef](#)]
43. Evans, I.M.; Hanft, R. *The Introduction of New Technologies*; ACOG Clinical Seminars, ACOG: Washington, DC, USA, 1997; p. 3.
44. Cohen, A.B.; Hanft, R.S. *Technology in American Health Care: Policy Directions for Effective Evaluation and Management*; University of Michigan Press: Ann Arbor, MI, USA, 2004; p. 480.
45. Britt, D.W.; Risinger, S.T.; Miller, V.; Mans, M.K.; Krivchenia, E.L.; Evans, M.I. Determinants of parental decisions after prenatal diagnosis of Down syndrome: Bringing in context. *Am. J. Med. Genet.* **2000**, *93*, 410–419. [[CrossRef](#)]
46. Britt, D.W.; Risinger, S.T.; Mans, M.; Evans, M.I. Devastation and relief: Conflicting meanings in discovering fetal anomalies. *Ultrasound Obstet. Gynecol.* **2002**, *20*, 1–5. [[CrossRef](#)]
47. Hui, L.; Bianchi, D.W. Fetal fraction and noninvasive prenatal testing: What clinicians need to know. *Prenat. Diagn.* **2020**, *40*, 155–163. [[CrossRef](#)]
48. Che, H.; Villela, D.; Dimitriadou, E.; Mellote, C.; Brison, N.; Neofytou, M.; Bogaert, K.V.D.; Tsuiko, O.; Devriendt, K.; Legius, E.; et al. Non-invasive prenatal diagnosis by genome-wide haplotyping of cell-free plasma DNA. *Genet. Med.* **2020**, *22*, 962–973. [[CrossRef](#)]
49. Vossaert, L.; Wang, Q.; Salman, R.; McCombs, A.K.; Patel, V.; Qu, C.; Mancini, M.A.; Edwards, D.P.; Malovannaya, A.; Liu, P.; et al. Validation study for single circulating trophoblast genetic testing as a form of noninvasive prenatal diagnosis. *Am. J. Hum. Genet.* **2019**, *105*, 1262–1273. [[CrossRef](#)]
50. Ravn, K.; Singh, R.; Hatt, L.; Kølvråa, M.; Schelde, P.; Vogel, I.; Uldbjer, N.; Hindkjær, J. The number of circulating fetal extravillous trophoblasts varies from gestational week 6 to 20. *Reprod. Sci.* **2020**, *27*, 2170–2174. [[CrossRef](#)]
51. Bianchi, D.W.; Simpson, J.L.; Jackson, L.G.; Elias, L.G.; Holzgreve, W.; Evans, M.I.; Dukes, K.A.; Sullivan, L.M.; Klinger, K.W.; Bischoff, F.Z.; et al. Fetal gender and aneuploidy detection using fetal cells in maternal blood: Analysis of NIFTY I data. National Institute of Child Health and Development Fetal Cell Isolation Study. *Prenat. Diagn.* **2002**, *22*, 609–615. [[CrossRef](#)] [[PubMed](#)]
52. Jou, H.J.; Chou, L.Y.; Chang, W.C.; Ho, H.C.; Zhang, W.T.; Ling, P.Y.; Tsai, K.H.; Chen, T.H.; Chen, S.H.; Lo, P.H.; et al. A novel automatic platform based on nanostructured microfluidic chip for isolating and identification of circulating tumor cells. *Micromachines* **2021**, *12*, 473. [[CrossRef](#)] [[PubMed](#)]
53. Huang, C.E.; Ma, G.C.; Jou, H.J.; Lin, W.H.; Lee, D.J.; Lin, Y.S.; Ginsberg, N.A.; Chen, H.F.; Chang, F.M.C.; Chen, W. Noninvasive prenatal diagnosis of fetal aneuploidy by circulating fetal nucleated red blood cells and extravillous trophoblasts using silicon-based nanostructured microfluidics. *Mol. Cytogenet.* **2017**, *2*, 44. [[CrossRef](#)] [[PubMed](#)]
54. Ma, G.C.; Lin, W.H.; Huang, C.E.; Chang, T.Y.; Liu, C.Y.; Yang, Y.J.; Lee, M.H.; Wu, W.J.; Chang, Y.S.; Chen, M. A silicon-based coral-like nanostructured microfluidics to isolate rare cells in human circulation: Validation by SK-BR-3 cancer cell line and its utility in fetal nucleated red blood cells. *Micromachines* **2019**, *10*, 132. [[CrossRef](#)]
55. Dondorp, W.; de Wert, G.; Bombard, Y.; Bianchi, D.W.; Bergmann, C.; Borry, P.; Chitty, L.S.; Fellmann, F.; Forzano, F.; Hall, A.; et al. Non-invasive prenatal testing for aneuploidy and beyond: Challenges of responsible innovation in prenatal screening. *Eur. J. Hum. Genet.* **2015**, *23*, 1438–1450. [[CrossRef](#)]