

Identifying Prenatal Alcohol Exposure and Children Affected by It: A Review of Biomarkers and Screening Tools

Julie A. Kable^{1,2} and Kenneth Lyons Jones³

¹Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia

²Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia

³Department of Pediatrics, University of California San Diego, La Jolla, California

Correspondence

Address correspondence concerning this article to Julie A. Kable, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, 12 Executive Park, Atlanta, GA 30322. Email: jkabl01@emory.edu

Disclosures

The authors declare no competing financial or nonfinancial interests.

Publisher's Note

Opinions expressed in contributed articles do not necessarily reflect the views of the National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health. The U.S. government does not endorse or favor any specific commercial product or commodity. Any trade or proprietary names appearing in *Alcohol Research: Current Reviews* are used only because they are considered essential in the context of the studies reported herein.

PURPOSE: Early identification of prenatal alcohol exposure (PAE) and of those in need of services resulting from this exposure is an important public health concern. This study reviewed the existing literature on potential biomarkers and screening tools of PAE and its impact.

SEARCH METHODS: Electronic databases were searched for articles published between January 1, 1996, and November 30, 2021, using the following search terms: (“fetal alcohol” or “prenatal alcohol” or “FASD” or “alcohol-related neurodevelopmental disorder” or “ARND” or “ND-PAE”) and (“screening” or “identification” or “biomarker”). Duplicate articles were electronically eliminated. Titles and abstracts were reviewed for appropriateness, and selected articles were retrieved for further analysis. Additional articles were added that were referenced in the reviewed articles or identified from expert knowledge. Information about the characteristics of the sample, the biomarker or screening tool, and the predictive validity outcome data were abstracted. A narrative analysis of the studies was then performed on the data.

SEARCH RESULTS: A total of 3,813 articles were initially identified, and 1,215 were removed as duplicates. Of the remaining articles, 182 were identified as being within the scope of the review based on title and abstract inspection, and 181 articles were successfully retrieved. Of these, additional articles were removed because they were preclinical (3), were descriptive only (13), included only self-report of PAE (42), included only mean group comparison (17), were additional duplicates (2), focused on cost analysis (9), missed predictive validity data (24), or for other reasons (23). The remaining articles ($n = 48$) were abstracted. An additional 13 manuscripts were identified from these articles, and two more from expert knowledge. A total of 63 articles contributed to the review.

DISCUSSION AND CONCLUSIONS: Biomarkers and screening tools of PAE and its impact fall short of ideal predictive validity characteristics. Higher specificity than sensitivity was found for many of the biomarkers and screening tools used to identify PAE and its impact, suggesting that current methods continue to under-identify the full range of individuals impacted by PAE. Exceptions to this were found in recent investigations using microRNAs related to growth and vascular development, proteomic changes associated with PAE, and combinations of markers estimating levels of various cytokines. Replications of these findings are needed across other samples to confirm the limited data available. Future research on biomarkers and screening tools should attend to feasibility and scalability of implementation. This article also recommends a systematic process of evaluation to improve early identification of individuals impacted by PAE so that harm reduction and rehabilitative care efforts can be implemented.

KEYWORDS: alcohol; prenatal alcohol; FASD; identification; biomarkers; fetal alcohol spectrum disorders; prenatal exposure delayed effects

Although the awareness of the negative impact of prenatal alcohol exposure (PAE) was already alluded to in ancient writings¹ and the impact of ethanol embryopathy in animal models was studied as early as 1910,² the conceptualization of a syndrome associated with PAE was not recognized within modern medicine until the mid-20th century.^{3,4} The syndrome or disorder was not uniformly accepted, however, and debates occurred within the field related to the operationalization of criteria for making a clinical diagnosis. In 1996, a group of scientists were brought together under the auspices of the Institute of Medicine (IOM) to delineate criteria for a diagnosis and a public health care plan for addressing the needs associated with the condition.⁵ This committee established the first consensus criteria for fetal alcohol syndrome (FAS) and recognized associated conditions, such as partial FAS (pFAS), alcohol-related birth defects (ARBD), and alcohol-related neurodevelopmental disorder (ARND). Various operational definitions of the IOM report's diagnostic guidelines have been used to make a clinical diagnosis.⁶⁻¹⁷ In all cases, these diagnostic formulations struggle with identifying infants negatively impacted by PAE because few tools are available for assessing early brain development. In addition, many of the diagnostic formulations require input from complex medical teams evaluating different domains of impact, which are costly and heavily constrained by the number of professionals qualified to carry out the assessments.

Estimates of the prevalence of prenatal alcohol-related disorders have varied dramatically over the years. In the initial IOM report, which reviewed several registries and clinic-based studies, the estimate of FAS was reported to be in the range of 0.5 to 3 cases per 1,000 births;⁵ however, more recent estimates have been much higher. A large consortium that estimated the prevalence of fetal alcohol spectrum disorders (FASD)—an umbrella term used to refer to a range of conditions (FAS, pFAS, ARBD, and ARND) associated with PAE—in four communities within the United States using active case ascertainment yielded a conservative estimate of 11.3 to 50 per 1,000 births¹⁸ and an even higher weighted prevalence estimate of 31 to 99 per 1,000 births. A review of more than 24 unique studies carried out throughout the world resulted in a prevalence estimate of 8 per 1,000 births with a 95% confidence interval of 5 to 12 per 1,000 births.¹⁹ Variations in the estimates are likely related to differences in diagnostic criteria used to estimate the prevalence of the disorder across studies, use of active versus passive surveillance methods, and regional variations in drinking patterns. Historically, documentation of PAE has been difficult to obtain due to unreliability of the self-report of women drinking in pregnancy and potential social stigma associated with acknowledging alcohol use in pregnancy that can result in underreporting of PAE.²⁰ The lack of recognition by various health professionals for the cluster of symptoms associated with the diagnosis of FASD also has contributed to under-recognition of those impacted by PAE.²¹

In anticipation of this problem, the IOM report outlined the need for biological markers of alcohol teratogenesis to help with resolving variations in case definitions.⁵ The term “biomarker” refers to a broad collection of medical signs that can be used to identify a disease and can be measured accurately and reliably.²² Biomarkers differ from medical symptoms, which are collected via patient report of their status and typically refer to biological measurements associated with the disease state. Biomarkers have the advantage of reducing ambiguity in patient reporting of symptoms but are only useful if they can validly predict a clinical endpoint—that is, if they can appropriately identify the disease state and avoid misclassification of individuals who do not have the condition. In the case of PAE, the clinical endpoint may be the identification of an alcohol-exposed pregnancy or of those negatively impacted by their exposure. Ideally, the identification would occur as early as possible during or after pregnancy to enhance opportunities for intervention. Identification during pregnancy could lead to harm reduction efforts, whereas early postnatal recognition of infants negatively impacted by PAE would increase the opportunities for access to habilitative care to optimize early brain development during phases of high neuroplasticity.²³ In addition to biomarkers, screening tools that sample symptoms of the disease state, or some combination of these, may be useful in identifying those negatively impacted by PAE. The development of innovative methods and tools that can be used to reduce the costly diagnostic assessment burden that constrains the identification of individuals in need of services are of particular value as such tools would allow for improved scalability and implementation in resource-poor areas of the world.

This review attempts to clarify potential advancements in the identification of biomarkers of PAE or its impact that could be used to improve early recognition of those adversely affected since the original IOM report's call for the development of biomarkers of alcohol-related teratogenesis. To this end, the authors conducted a review of the literature on the predictive validity of biomarkers or screening tools for identification of PAE or FASD and performed a narrative analysis of the findings.

Search Methods

Studies were considered for review if the article was published or available online between January 1, 1996, the first day of the IOM report publication year, and November 30, 2021. The target population consisted of individuals of any age who had been diagnosed with PAE or with a clinical disorder associated with PAE (i.e., FAS, pFAS, ARND, ARBD, and neurobehavioral disorder associated with prenatal alcohol exposure [ND-PAE]).²⁴

In addition, the article's focus had to include screening or identification of PAE or one of the clinical disorders associated with PAE. The article also had to include empirical data related

to the screening or identification procedures and provide some aspect of the biomarker's predictive characteristics. Predictive validity characteristics evaluated in each study included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and area under the curve (AUC). Sensitivity refers to the probability that the test is positive when the condition is present. Specificity refers to the probability that the test is negative when the condition is not present. PPV refers to the probability that the condition is present when the test is positive. NPV refers to the probability that the condition is not present when the test is negative. Accuracy refers to the overall probability that the case is correctly classified from the test. Criterion descriptors for the predictive values are as follows: 90–100%, Excellent; 80–89%, Good; 70–79%, Fair; and below 70%, Poor. Finally, AUC is derived from creating receiver operating curves by plotting the true positive rate (sensitivity) relative to the false positive rate (1-specificity). The AUC references the area on the graph created by the regression line relative to the chance rate of prediction. Values of 1 would indicate perfect condition, and values of 0.50 would indicate chance prediction using a binary (yes/no) model.

Definitions for the first five predictive validity characteristics and formulas for computing them are outlined in Figure 1, a confusion matrix that illustrates the classic prediction modeling used when comparing a test's ability to identify a given state or condition. The confusion matrix is a contingency table that presents the frequency of individuals categorized across two dimensions, the actual true state of whether or not an individual has a disease or condition, and the predicted state derived from

the results of the testing indicating the presence of the disease or not.

To identify studies, the following electronic databases were searched: PsycInfo, PubMed, Medline, Web of Science, ERIC, and the Cochrane Central Register of Control Trials. Search terms used were ["fetal alcohol" or "prenatal alcohol" or "FASD" or "alcohol-related neurodevelopmental disorder" or "ARND" or "ND-PAE"] and ["screening" or "identification" or "biomarker"]. Document type was limited to "articles," but no language restrictions were placed on the initial search. Despite extensive work in animal models of PAE on various promising biomarkers, only articles using humans were selected as the focus of this study was to analyze the current knowledge of potential tools that could be used to identify people affected by PAE. Preclinical biomarker methodologies still need translation into human populations to effectively evaluate their predictive characteristics.

References were then merged into Endnote X9.3.1 and screened for duplicates. The remaining studies were then reviewed to eliminate nonempirical studies (i.e., reviews or editorial articles) and those involving training of professionals to screen. Articles were also excluded if they established group differences without analyzing the predictive validity of the outcome or were descriptive of PAE in a given population. While establishing group differences may be a first step in establishing the utility of a biomarker or screening tool, such differences do not establish a tool's predictive utility. IQ tests are a classic example of tools that consistently demonstrate group differences between PAE groups relative to community samples without exposure;²⁵ however, they have little predictive utility when used independently as a result of the wide range of outcomes seen in

		True State		
		Positive	Negative	
Test	Positive	True Positive A	False Positive C	PPV $(\text{Sensitivity} \times \text{Prevalence}) / (\text{Sensitivity} \times \text{Prevalence} + (1 - \text{Specificity}) \times (1 - \text{Prevalence}))$
	Negative	False Negative B	True Negative D	NPV $(\text{Specificity} \times (1 - \text{Prevalence})) / (1 - \text{Sensitivity}) \times \text{Prevalence} + \text{Specificity} \times (1 - \text{Prevalence})$
		Sensitivity $A / (A + B)$	Specificity $D / (C + D)$	Accuracy $\text{Sensitivity} \times \text{Prevalence} + \text{Specificity} \times (1 - \text{Prevalence})$

Figure 1. Confusion matrix. The confusion matrix provides definitions of the various predictive validity terms within a contingency table where cases are plotted relative to the prediction variable and the designated "true state." True state refers to whether the individual has a disease or condition (positive) or does not have a disease or condition (negative), and the test reflects the outcome of the criterion used to indicate a positive or negative prediction of disease state. Sensitivity refers to the probability that the test is positive when the condition is present. Specificity refers to the probability that the test is negative when the condition is not present. PPV refers to the probability that the condition is present when the test is positive. NPV refers to the probability that the condition is not present when the test is negative. Accuracy refers to the overall probability that the case is correctly classified from the test. Note: NPV, negative predictive value; PPV, positive predictive value.

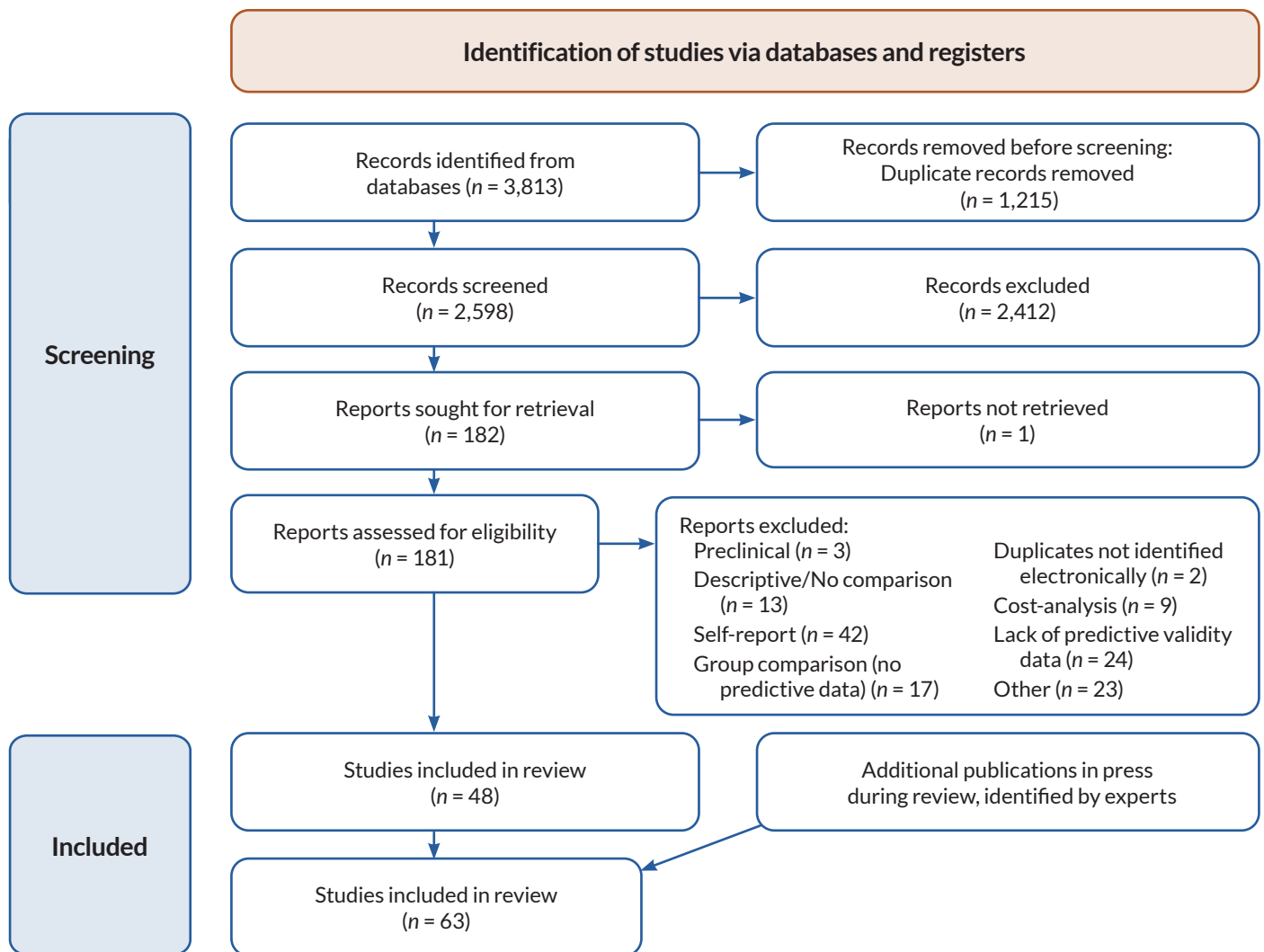


Figure 2. Flow diagram of the steps in the screening process for this review.

individuals with PAE and its associated overlap with comparison samples. A flow diagram (Figure 2) outlines the various steps in screening the articles and the number of articles at each step.

Search Results

A total of 3,813 articles were initially captured by the search, and 1,215 were identified as duplicates. Article titles and abstracts were then screened for inclusion, and an additional 2,412 were eliminated, leaving 181 full articles that were retrieved. One article could not be retrieved. The full articles were reviewed for appropriateness, and 133 articles were excluded for the following reasons: three were preclinical, 13 were descriptive only, 42 related to predictive utility of self-report methods of PAE, 17 were identified as group comparison studies, two were additional duplicates not identified electronically, nine were related to cost analysis, and 24 after further review did not have predictive data. This left 48 articles; however, upon further

review, 13 additional articles were identified that were not retrieved by the search. Moreover, two additional articles were identified based on expert knowledge. This resulted in 63 articles included in the review.

Biomarkers and screening tools were categorized as predicting prenatal exposure status or alcohol-related teratogenesis in the offspring. Appendix 1 provides details on the articles that involved biomarker predictors of PAE status, and Appendix 2 provides details on biomarker predictors of FASD and associated symptoms. Both appendices list the articles in alphabetical order by the first author's last name as many involve the evaluation of several biomarkers and predictors within one study. Appendix 3 provides details on other screening tool predictors of FASD and associated symptoms, including craniofacial features, neurophysiological responses, neuroimaging analyses, questionnaire responses, and various test batteries assessing performance. As typically only one screening tool was evaluated within a study, Appendix 3 groups studies by screening tool category and then lists studies alphabetically.

Predictive validity information was obtained from information explicitly stated in the text or tables or was computed from information regarding cell sizes in the predictive validity tables provided in the article or as described in the text. Computations were performed using MedCalc software for diagnostic test evaluation (MedCalc Software Ltd, Ostend, Belgium). Predictive validity values are presented as percentages with the exception of AUC values, which were reported in proportions of accurate diagnostic classification with values of 0 to 1.00.

The sensitivity, specificity, accuracy, and AUC values were plotted on radial curves for each type of biomarker, with each type of predictive characteristic color-coded (see Figure 3). AUC values were multiplied by 100 to facilitate plotting them on the same curves as the other predictive values. The obtained values for each of the validity characteristics were provided for each unique outcome of the study. For studies that compared the biomarker response to common outcomes defined differently (e.g., self-report using different assessment tools), only the obtained values reflecting the least and greatest value were included to reflect the range of validity. Radial curves plot individual values of these predictive parameters along a curve with increasing number of indicators smoothing out until the curve is circular. The strength of the prediction is reflected along the radius of the circle so that values in the outer region reflect increased predictive validity and those in the inner region reflect lower levels of predictive validity. Radial curves allow for a quick visual analysis of each of the predictive characteristics for each type of biomarker or screening tool and the variation across the findings. Curves with more points along the outer ring with less deviance inward reflect increased predictive status and uniformity in the prediction.

Biomarkers

Biomarkers of PAE were derived from various biological samples obtained from mothers, including blood (plasma and dried blood spots), urine, hair, and fingernail clippings. Sources of biomarkers evaluated in the infant included blood (plasma and dried blood spots) and meconium. Additional biomarkers of PAE or its effects were obtained from placental tissue and the umbilical cord. Biomarkers were evaluated against group status determined from maternal self-report of alcohol consumption and the offspring's FASD symptomatology or diagnosis.

One group of biomarkers evaluated included fatty acid ethyl esters (FAEE) derived from hair or meconium. FAEE are metabolites of ethanol and provide a long-term estimate of alcohol consumption over the course of a pregnancy. They were analyzed either in a collective grouping of FAEE or individually (i.e., ethyl stearate, ethyl linoleate); in total, 30 obtained values or point estimates of predictive validity were provided across 12 studies.²⁶⁻³⁷ In three additional studies, FAEE were used as the outcome to assess other biomarker predictors.³⁸⁻⁴⁰ The radial graph of the predictive characteristics of FAEE in combination or

separately (see Figure 3A) suggests that their specificity (range, 43%–100%; median, 83%) as biomarkers is significantly better than their sensitivity (range, 4%–100%; median, 65%); overall accuracy estimates fell in the poor to fair range (range, 62%–79%; median, 68%). Estimates of the AUC values were variable, ranging from poor to excellent (range, 0.52–0.93; median, 0.71). There was no clear pattern that a summation of several FAEE or any one FAEE provided better prediction.

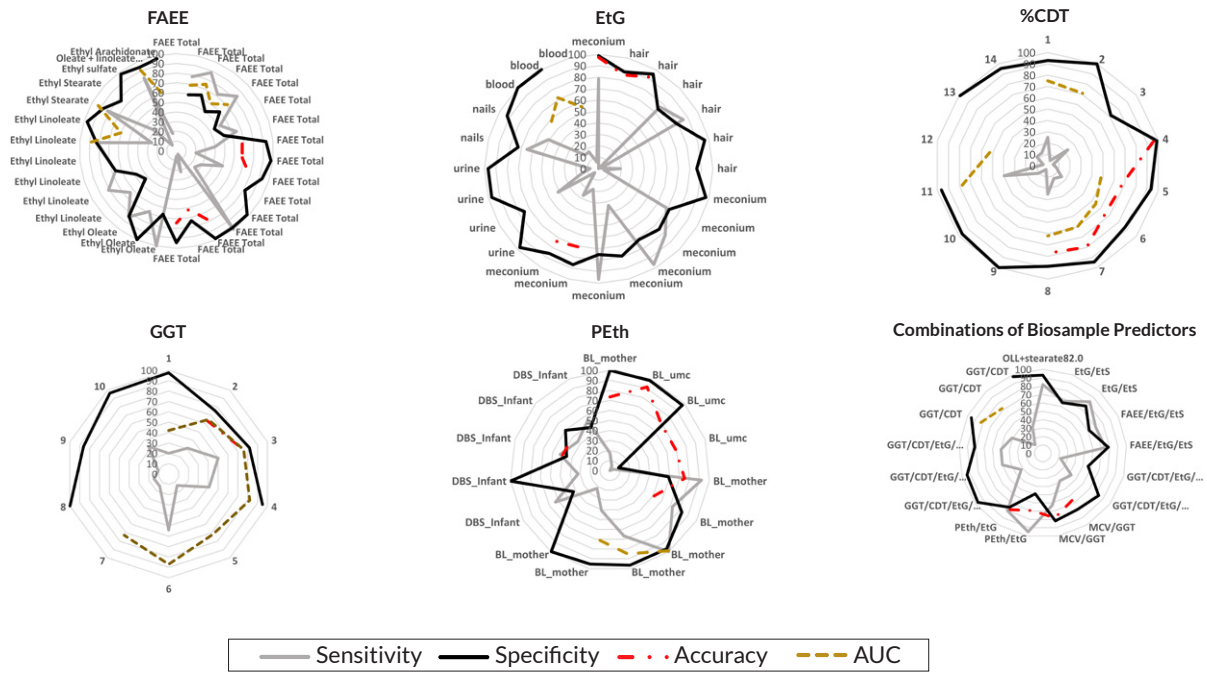
Other biomarkers assessed included gamma-glutamyltransferase (GGT),^{35,41-46} carbohydrate-deficient transferrin (CDT),^{38,41-46} ethyl glucuronide (EtG),^{30,31,35,38-41,43,44,47-52} ethyl sulfate (EtS),^{31,35,41} and mean corpuscular volume (MCV).⁴⁵ GGT, CDT, and MCV provide an indirect assessment of the impact of heavy and chronic alcohol use on the mother's metabolic functioning. Estimates of GGT can be obtained from plasma, urine, and hair, whereas CDT and MCV estimates are only obtained from plasma. EtG and EtS are metabolites of ethanol that are present in hair, meconium, urine, and nails. Predictive validity information was found for seven studies using GGT (10 point estimates), seven studies using CDT (13 point estimates), and one study using MCV (three point estimates). Fifteen studies with 24 point estimates were identified for EtG. Three studies evaluated EtS,^{31,35,41} but only two provide estimates of EtS alone,^{35,41} whereas one study evaluated EtS in combination with EtG.³¹ Consistently, these biomarkers provided fair to excellent specificity—EtG (range, 71%–100%; median, 87%); EtS (range, 97%–100%; median, 98%); CDT (range, 71%–100%; median, 95%); GGT (range, 71%–100%; median, 95%); and MCV (both values 100)—but exceptionally poor sensitivity—EtG (range, 0%–97%; median, 23%); EtS (range, 7%–15%; median, 7%); CDT (range, 5%–40%; median, 13%); GGT (range, 11%–50%; median, 25%); and MCV (values of 15 and 20).

One study evaluated postnatal serum levels of insulin-like growth factor-II (IGF-II) as predictors of FASD status in children or youth who either had a history of meconium FAEE levels above 2 nmol/g or had been adopted from Eastern European countries with confirmed PAE (two point estimates).²⁷ The participants were assessed for IGF-II levels below the 5th percentile. IGF-II levels below the 5th percentile had excellent specificity (99% and 100%, respectively) for predicting FASD status, but very poor sensitivity (13% and 39%, respectively) and overall accuracy (24% and 47%, respectively).

One study provided limited information on aspartate aminotransferase (AST) and alanine aminotransferase (ALT), which are both markers of impaired liver functioning, as biomarkers of PAE.⁴⁶ Only AUC values were provided, and these were poor (0.47 and 0.54, respectively).

Phosphatidylethanol (PEth) is a more recent biomarker of ethanol metabolism that has been evaluated in maternal and infant plasma and dried blood spots.^{35,41,52-55} Six different studies found considerable variability in the predictive characteristics of PEth depending on the source of the PEth. Assays of maternal blood as well as plasma from the umbilical cord yielded a wide

3A



3B

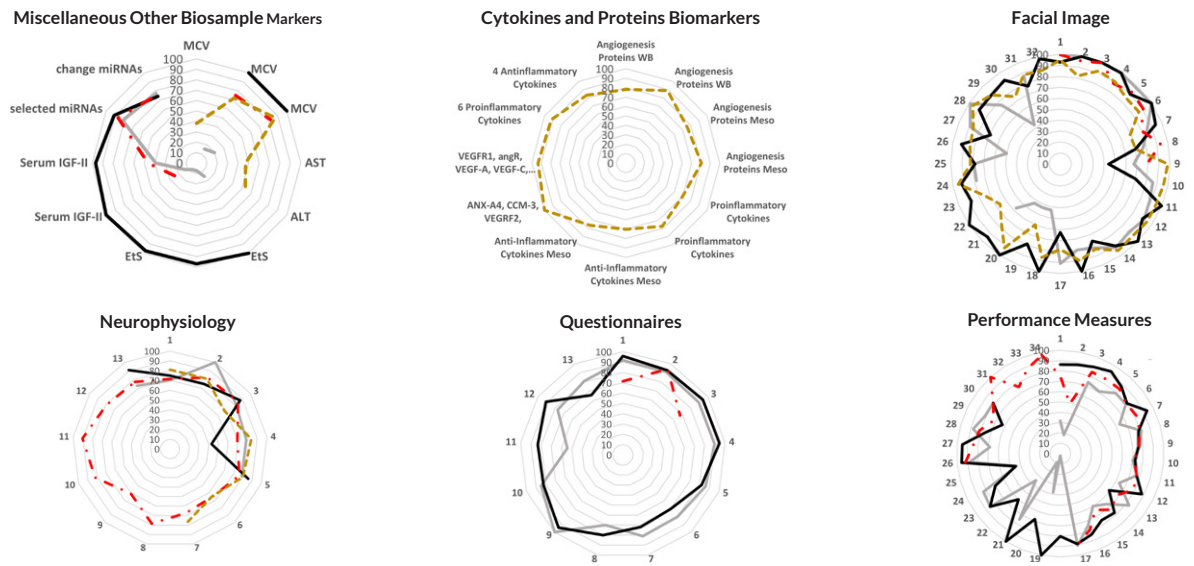


Figure 3. Radial curves of PAE biomarkers (A) and of biomarkers and screening tools for PAE and its impact (B). Radial graphs indicate the specificity (gray curves), sensitivity (black curves), accuracy (red dotted curves), and area under the curve (AUC) values (gold curves) relative to the criterion evaluated in the study. Point estimates or the obtained values of the validity characteristics were provided for each unique outcome of the study. For studies that compared the biomarkers' response to common outcomes defined differently (e.g., self-report using different assessment tools) only the point estimates reflecting the least and greatest value were included to reflect the range of validity. AUC values were multiplied by 100 to facilitate plotting them on the same curves as the other predictive values. The radial graph plots the various findings above curves with increasing prediction (0–100). Radial curves allow for a quick visual analysis of each of the predictive characteristics for each type of biomarker or screening tool and the variation across the findings. Greater numbers of findings displayed in a graph result in smoothing of the curve. The strength of the prediction is reflected along the radius of the circle so that values in the outer region reflect greater predictive validity and those in the inner region reflect lower levels of predictive validity. Curves with more points along the outer ring with less deviance inward reflect increased predictive status and uniformity in the prediction. Separate colored lines are used to connect the points along with curve for each of the predictive characteristics. Criterion descriptors for the values plotted above are as follows: 90–100, Excellent; 80–89, Good; 70–79, Fair; and below 70, Poor.

Note: ALT, alanine aminotransferase; angR, angR protein; ANX-A4, annexin-A4; AST, aspartate aminotransferase; AUC, area under the curve; BL, blood level; CCM-3, cerebral cavernous malformation 3 (a protein); CDT, carbohydrate-deficient transferrin; a protein; DBS, dried blood spots; EtG, ethyl glucuronide; EtS, ethyl sulfate; FAEe, fatty acid ethyl esters; GGT, gamma-glutamyltransferase; IGF-II, insulin-like growth factor-II; MCV, mean corpuscular volume; miRNAs, micro RNAs; NPV, negative predictive value; OLL, oleate + linoleate + linolenate; PEth, phosphatidylethanol; PPV, positive predictive value; umc, umbilical cord; VEGF, vascular endothelial growth factor; VEGFR, vascular endothelial growth factor receptor; WB, Western Blotting Procedures.

range of specificity (range, 9%–100%; median, 96%), sensitivity (range, 0%–100%; median, 22%), and overall accuracy (range, 51%–91%; median, 71%). Tests of dried blood spots taken from infants also had variability in their predictive characteristics but were generally not as good as maternal blood and plasma obtained from the umbilical cord—specificity (range, 42%–100%; median, 95%); sensitivity (range, 32%–63%; median, 52%); and overall accuracy (range, 48%–50%; median, 50%).

Collectively, these results regarding the validity of biomarkers for predicting PAE status suggest that a positive response was not very effective in identifying the full range of individuals who self-reported prenatal alcohol use and missed many affected individuals. This was also true of the studies evaluating the predictive modeling of the impact of PAE (see Appendix 2). Combining biomarkers did not result in substantial improvements in the predictive characteristics (see Figure 3A, bottom right panel). As has been observed in other biomarker analyses, there appeared to be a trade-off such that as sensitivity of combined biomarkers increased compared with single biomarker predictors, specificity was reduced.

A promising biomarker with limited predictive data reported in one study was proteins and cytokines found in the placenta.⁵⁶ Specifically, proteins that influence angiogenesis as well as pro-inflammatory and anti-inflammatory cytokines were evaluated in a group with a history of PAE. The study only provided information on AUC, which reflects the integration of sensitivity and specificity characteristics; however, these data were in the fair to excellent range (range, 0.70–1.00; median, 0.79). In contrast to previous biomarker data, integration of different predictors resulted in improved prediction. Combined analysis of the levels of three proteins (i.e., ANX-A4, CCM-3, and VEGFR2) yielded an AUC of 1.00, and a combined analysis of another six proteins (VEGFR1, angR, VEGF-A, VEGF-C, VEGF-D, and beta-fibroblast growth factor) resulted in an AUC of 0.94. Combined cytokine levels also had good to excellent AUC values, with six pro-inflammatory cytokines (IL-1-beta, IL-2, IL-8, IL-12p70, interferon-gamma, and tumor-necrosis factor alpha) yielding an AUC value of 0.92 and four anti-inflammatory cytokines (IL-4, IL-6, IL-10, and IL-13) resulting in an AUC value of 0.83.⁵⁶

Finally, circulating microRNAs (miRNAs) in maternal blood, which reflect epigenetic changes in response to PAE, have been explored as a potential biomarker in a sample of Ukrainian mother-infant dyads.⁵⁷ Levels of miRNAs were compared among pregnant women without PAE; pregnant women with heavy PAE whose children were impacted; and pregnant women with heavy PAE whose children were not impacted in either growth, dysmorphology, or brain development. Heavy PAE was defined as weekly heavy episodic or binge drinking (i.e., five or more standard drinks), five or more episodes of three to four standard drinks, or 10 episodes of one to two standard drinks. Impact of PAE on the offspring was assessed by trained physicians who completed a dysmorphology assessment and by psychologists who completed a neurodevelopmental evaluation with the child.

Several miRNAs ($n = 21$) were identified as differing between the exposed-affected group and both other groups, and a random forest analysis was used to predict group membership while controlling for other group differences (i.e., maternal smoking). Seven of the top 10 variables retained in the initial predictive model were miRNAs. The most common miRNAs identified were likely to influence downstream pathways related to fetal and placental growth. Specificity was excellent (91%) and sensitivity (82%) was good for miRNA levels obtained in pregnancy; however, both specificity (74%) and sensitivity (77%) were only fair for changes in the miRNA levels over the course of the pregnancy. Although this was only one study, the findings suggest that assessments of levels of specific miRNAs obtained in pregnancy may improve sensitivity in predicting PAE-related outcome compared with other biomarkers that could be obtained in pregnancy.

Screening Tools

Screening tools were divided into five types of assessments, including facial features, neurophysiological responses in infants and older children, neuroimaging, questionnaire responses, and performance measures (see Appendix 3). In some cases, combinations of facial data and performance measures were used in predictive modeling; these are included in the performance measure section of Appendix 3.

Facial features

Eight studies have explored facial features as key predictors of an FASD-related diagnosis using in-person measurements and two-dimensional (2D) and three-dimensional (3D) photographs.^{58–65} Specificity values were variable, ranging from poor to excellent, with only a couple of studies reporting levels in the fair to poor range (range, 43%–100%; median, 86%). Sensitivity levels also were in the good to fair range (overall range, 43%–100%; median, 92%), with the exception of one study where sensitivity using the facial analysis software of 2D pictures was in the poor range. Accuracy for prediction was typically in the fair to good range (range, 79%–100%; median, 93%). Advancing technology from in-person measurement to 3D computerized configural methods did not necessarily result in improved predictive characteristics, but comparisons are complicated because samples were from different countries (i.e., United States, South Africa, Germany, and Finland), and different methods were used for defining the outcome (variations of FAS and FASD, heavy alcohol-exposed) and reporting predictive results.

More recently, one study evaluated the use of a schema that coded alterations to ocular development to differentiate individuals with a clinical diagnosis of FASD.⁶⁶ The coding schema captured elements of visual acuity, refraction, strabismus/binocular function, and ocular structural abnormalities, with each area being coded from 1 to 4. Cut-off values of the total score (10 and 9) were evaluated relative to healthy controls; children with attention-deficit/hyperactivity disorder (ADHD);

children who were born prematurely (moderate to late); and children with Silver-Russell syndrome, a genetic condition with growth impairment and neurodevelopmental compromise.⁶⁷ Similar to attempts to capture facial features, specificity was good to excellent (88%–100%), but sensitivity was poor (43%–57%). AUC estimates were variable, ranging from 0.60 to 0.92, with the higher estimate reflecting comparisons to healthy controls.

Infant neurophysiology

Early identification of alcohol-related brain impairment has been attempted using indices of infant neurophysiological responses, including eye-blink conditioning⁶⁸ and cardiac orienting response (COR).^{69,70} These procedures use physiological responses in the context of a learning paradigm that can be implemented with infants. For eye-blink conditioning, classical conditioning is used where an unconditioned stimulus (i.e., puff of air) that elicits a reflexive eye blink is paired with a conditioned stimulus (i.e., auditory tone or picture) over repeated trials. After many pairings, the conditioned stimulus is then able to elicit the eye-blink response. Rate of learning is assessed by the percentage of pairing trials of the conditioned stimulus with the unconditioned stimulus needed before the eye blink is elicited by the conditioned stimulus in the absence of the unconditioned stimulus. In the case of COR, heart rate responses are monitored while stimuli (i.e., auditory tone or picture) are presented over several trials, referred to as habituation trials, and then after presenting novel but similar stimuli over several trials (dishabituation trials). Heart rate typically decelerates in response to novel information and returns to baseline over the course of several habituation trials; it decelerates again in response to the second novel stimulus. The magnitude of the deceleration in the first three habituation trials is believed to reflect the infant's encoding of stimuli, whereas the magnitude of the first three dishabituation trials reflects the infant's ability to differentiate the first and second related stimuli, indexing early memory functioning. These methods are advantageous as standardized early assessments of cognitive functioning often are not adequate in assessing alcohol-related brain impairment.

Eye-blink conditioning was reported in one study that provided data for its predictive utility relative to FAS and to a broader spectrum of individuals with heavy PAE, defined as averaging at least 1.0 oz absolute alcohol per day or \geq five standard drinks per occasion in the first trimester of pregnancy; and a group defined as having FASD.⁶⁸ Eye-blink conditioning had a sensitivity of 100% for FAS prediction, but this fell to 70% for prediction of a broader spectrum of heavy PAE and FASD. Specificity was comparable for both predictive models at 75%. Overall accuracy was 82% for predicting FAS and 72% for predicting heavy PAE/FASD. The PPV value was 87% for heavy PAE/FASD and 63% for FAS alone, and NPV was 51% for predicting heavy PAE/FASD and 100% for FAS alone.

Findings for COR were not reported in terms of sensitivity, specificity, and overall accuracy but were reported in terms of PPV, NPV, and AUC values in two different articles using overlapping samples of Ukrainian mother-infant dyads.^{69,70} Using the key features of COR (i.e., speed of the response, average trough), a PPV of 82%, an NPV of 62%, and an AUC value of 0.81 were reported in one of the studies for predicting neurodevelopmental impairment at 12 months.⁷⁰ Only small incremental gains were obtained when including maternal drinking information in the model. In the second study, an index score derived from the visual COR data had an AUC value of 0.77 for predicting later preschool FASD status.⁶⁹ These results suggest that early neurophysiological responses may be useful in improving identification of individuals with neurodevelopmental impairment in infancy, which has often been a key factor limiting early diagnosis.

Neurophysiology with older children

Neurophysiological responses assessed in older children have included auditory evoked potentials and eye-tracking or saccadic eye movements. One study evaluated auditory evoked potentials, which assess the time it takes for a signal to travel along the auditory nerve track in response to sound stimuli.⁷¹ Auditory evoked potentials by themselves had fair sensitivity (79%) and poor specificity (43%) and overall accuracy (61%). However, when various indices of P300 responses were combined (e.g., latency, magnitude), increased differentiation of individuals with FASD from individuals with Down syndrome was found (sensitivity, 79%; specificity, 86%; and overall accuracy, 82%).

Eye-tracking movements also have been used to identify children impacted by PAE.⁷² Two studies provided data regarding predictive validity of eye-tracking measures in individuals with FASD.^{73,74} Accuracy ratings ranged from poor (65%) to excellent (90%). Combining eye-tracking information with data obtained from diffusion tensor imaging and neurobehavioral testing resulted in improved accuracy in one study (range of 65%–76% improved to 85%).⁷³ Eye-tracking movements also have been used to predict the impact of other neurodevelopmental disorders,^{75,76} suggesting the importance of studies that attempt to establish differential predictive validity for the effects of PAE relative to other neurodevelopmental disorders (e.g., autism). This likely is also true of the infant neurophysiological measures (i.e., COR and eye-blink conditioning), which also have been used to determine mean group differences between other clinical groups and typically developing controls.^{77,78}

Neuroimaging

Three neuroimaging studies provided predictive data for the impact of PAE.^{73,79,80} Using weighted volumetric scores of specific brain regions, specificity was good (88%), but sensitivity was still in the poor range (64%).⁸⁰ The combination of four key features of diffusion tensor imaging also provided relatively

poor accuracy (67%) in predicting an FASD diagnosis.⁷³ Excellent specificity (95%) was reported for measurement of the “hook” area of the corpus callosum, but sensitivity of this measurement was poor (52%), suggesting that this method did not identify those impacted by PAE at better than chance levels.⁷⁹ This suggests that, like other biomarker prediction of PAE and PAE impact, prediction based on neuroimaging findings provides a clear signal of PAE or its impact, but is not sufficiently sensitive to capture the range of impact commonly seen in individuals exposed to alcohol.

Parent questionnaire measures

Six identified studies reported predictive characteristics of caregiver or provider responses to a questionnaire in identifying children with alcohol exposure or FASD.⁸¹⁻⁸⁶ Parental responses to questionnaires developed specifically for identifying children impacted by PAE or standardized measures used to flag aspects of alcohol teratogenesis typically had good to excellent specificity (overall range, 66%–96%; median, 83%); only one study using subsets of items from the Child Behavior Checklist yielded sensitivity in the poor to fair range.⁸² Sensitivity reported in these studies was poor to excellent (range, 54%–100%, median, 85%), with the lowest sensitivity reported in a study attempting to differentiate only pFAS in one analysis (54%).⁸⁵ Relatively few studies reported overall accuracy rates, which ranged from poor to excellent (range, 68%–94%; median, 71%). The wide range in predictive characteristics of these types of data was dependent on the definition of the predictor (PAE, pFAS, FAS, or FASD) and the comparison group used—typical healthy controls or controls with ADHD. Incomplete evaluation of those who screened negative also may have overinflated estimates in one study of the predictive characteristics as this method fails to include the possibility of false negatives in the screening process.⁸⁴

Child performance measures

Nine studies identified predictive characteristics of child performance measures and combinations of performance measures and other indicators of PAE or FASD.⁸⁷⁻⁹⁵ These ranged from quick screening tests to complex neurobehavioral batteries in isolation or in combination with dysmorphology information. Of these nine studies, one assessed the predictive characteristics of motor assessments,⁹² whereas another two studies looked at aspects of narrative speech only.⁹⁴⁻⁹⁵ Specificity ratings for all nine studies ranged from poor (45%) to excellent (100%), and sensitivity ratings ranged from poor (2%) to excellent (100%). Overall accuracy in these studies also ranged from poor (49%) to excellent (100%). Two of the nine studies compared individuals with PAE to both typical healthy control groups and to other clinical groups separately or in combination with the healthy control group.^{89,90}

Discussion and Conclusions

Identifying children who have been prenatally exposed to alcohol or, more importantly, have been negatively impacted by their exposure continues to be an important area of investigation. Although a range of biomarkers and screening tools have been explored, there is no agreed-upon procedure or method that provides excellent sensitivity, specificity, and overall accuracy, suggesting the need for continued research. A general theme found in the existing literature is higher specificity than sensitivity for many of the biomarkers and screening tools used to identify PAE and its impact. This means that although researchers and clinicians often have confidence when they identify PAE or its impact, they struggle with capturing the full range of individuals impacted. Exceptions to this were found in recent investigations of biomarkers of PAE using miRNAs related to growth and vascular development,⁵⁷ proteomic changes associated with PAE,⁵⁶ and combinations of markers estimating levels of various cytokines.⁵⁶ However, replications of these findings across other samples are needed to confirm the limited data currently available on the predictive characteristics of these biomarkers.

For predicting the outcomes of alcohol teratogenesis, facial features operationalized using varying methods (i.e., in person, 2D, or 3D) provided relatively high sensitivity, specificity, and accuracy, but a few point estimates were less effective. Neurophysiological responses assessed in infancy and later childhood were able to differentiate individuals impacted by PAE, but the upper limits of prediction were in the fair to good range. Moreover, there was some indication that these responses were better at defining pFAS/FAS rather than the full spectrum of FASD, including heavy PAE. Neuroimaging methods, including volumetric and diffusion tensor imaging, also had high specificity but poor sensitivity, similar to biomarkers of PAE alone. Parent and professional responses to questionnaires had both good sensitivity and specificity, with the exception of one comparison that attempted to discriminate specific subgroups of FASD. This increased sensitivity relative to other biomarkers and screening tools may be biased by the fact that all studies in this area involved clinical FASD samples, which may reflect shared variance associated with the parent seeking treatment for the child. Replications in prospective cohorts of exposure may be helpful in clarifying this potential bias in predictive validity. Child performance measures had varying ranges of success in predicting those impacted by PAE, which seemed to vary as a function of inclusion of other biomarkers and the nature of the comparison sample utilized in the prediction.

Limitations in the Existing Literature

The definition of the criterion to be predicted was problematic across studies. Maternal report of PAE or heavy PAE was operationalized using multiple different methods that were

integrated in different ways (e.g., summed, any positive response, principal component analysis of several responses). Moreover, results appeared to vary as function of the context in which the maternal self-report was collected. In one study, maternal self-report of PAE was higher than PAE confirmed using biomarker data.⁵² In another study in the context of a health care environment, however, estimates of PAE using these methods were in the opposite direction.⁹⁶ Even in studies of FAEE levels that were conducted in the same hospital setting where participants were assured of confidentiality, FAEE levels were dramatically higher when they were sampled from de-identified meconium, which did not require maternal consent, than when informed consent from the mother was needed.⁹⁷ Mothers with the heaviest prenatal alcohol use were more likely to self-select out of the study,⁹⁸ most likely in response to the stigma associated with PAE.⁹⁹

A number of studies used other biomarkers to validate a novel biomarker. Convergent validity is useful in verifying the validity of the novel biomarker but limits the window of detection between biomarkers; moreover, threshold or cutoff values used to signal a positive test also varied. Often biomarkers reflect severe alcohol use disorder as they are indicators of damage to organs (e.g., liver) over a prolonged period; however, these methods often failed to capture the full range of FASD or PAE that can have adverse impact on a developing fetus. Other biomarkers are byproducts of the metabolism of alcohol and have limited windows for detecting PAE. For each biomarker, other factors also may reduce the validity of their prediction, including personal care and hygiene (e.g., corruption from chemicals used in hair and nail care), other foods that may produce alcohol metabolites during decomposition,⁹⁷ and willingness of the mother to provide the biological sample. Some investigators have opted to use a combined approach, although costly, to predict PAE status^{35,45,53} to compensate for the individual weaknesses or limitations of any one method of identification of PAE.

Many studies used an FAS or FASD diagnosis as the outcome, but diagnostic formulations used in the field vary considerably, and evidence suggests that the degree of agreement across methods is low.¹² The development of a consensus diagnostic formulation for individuals with FASD would be helpful in reducing error variance associated with the diagnostic formulations. As mentioned previously related to parental questionnaires as screening tools, use of clinical samples also is biased because it selects for individuals who sought care for the treatment of the child. This can result in circularity in defining the screening tool as the predictor when the screening tool may be drawn from the same construct domain or type of test used to categorize or diagnose the clinical group. Implementation of screening approaches across multiple samples—including both clinical and prospective cohorts of PAE from diverse populations

that vary in ethnic, geographic, and cultural backgrounds—may help with eliminating these biases.

Another limitation of some studies was that they provided predictive estimates but failed to sample the criterion within the entire pool of individuals screened.^{59,84} This approach occurred in larger screening cohorts where individuals who screened negative were not sampled further and were assumed to be true negatives. These assumptions may result in overestimation of the predictive characteristics of the biomarker or screening tool.

Sensitivity and specificity characteristics are independent of the prevalence of the condition under investigation (e.g., PAE), but accuracy, PPV, and NPV are influenced by the rate of PAE or individuals impacted by PAE in a study's sample (see Figure 1 for computational formulas). Considerable variation existed across studies in the ratios of affected and nonaffected individuals in the sample. In many studies, both groups were comparable in size, which results in an estimate of the predictive characteristics under circumstances where the prevalence of the condition in the sample is substantially higher than the rate anticipated in the general population. Changes in the sensitivity and specificity of a biomarker if the prevalence of the condition deviates from 50% can result in reduced validity of estimates of the overall accuracy of a biomarker or screening tool.¹⁰⁰ This suggests that the accuracy ratings commonly found for biomarkers of PAE and its impact may be overweighted by their high specificity and that these biomarkers are less predictive in real-world settings where the prevalence has been estimated to fall between 5 to 50 per 1,000 children.^{18,19} Implementation of biomarkers or screening tools in clinical trials in the context in which they are intended to be used may help to evaluate the true accuracy of these tools.

The studies surveyed also differed in comparison samples used, with some studies including typical healthy controls and others attempting to differentiate offspring with PAE relative to other clinical groups who might present for diagnosis. Estimates of predictive validity of biomarkers or screening tools relative to typical healthy controls are often higher than those found when using a clinical comparison group. However, the latter approach provides a better estimate of the usefulness of a biomarker or screening tool to clinicians asked to determine if a given child has been impacted by their PAE. In evaluating biomarkers or screening tools, researchers should consider a tiered approach with a first evaluation relative to typical controls, followed by evaluation relative to other clinical groups to improve understanding of the clinical utility of the biomarker or screening tool. The final tier would then involve an actual clinical trial of the clinical utility of the biomarker or screening tool and an assessment of where it fits within a clinical diagnostic algorithm—that is, whether it functions more as a screener that can flag the need for other diagnostic assessments or as an actual diagnostic tool, indicating its high concordance with the clinical endpoint.

Finally, the scalability of a biomarker or screening tool is also important to consider. The financial cost of the assay or test and the expertise needed to carry out an assessment can

dramatically limit the utility of a given biomarker or screening tool, particularly in countries with low resources. The gold standard for diagnosis is a multidisciplinary team assessment that includes at a minimum a physician who can assess alcohol-related dysmorphology and a psychologist who can assess neurobehavioral impairment. Even if variations in diagnostic criteria utilized among existing clinics are resolved, this method of identification in no way can meet the needs of those impacted by PAE given the recent prevalence estimates. This is true in countries with considerable resources as well as in those with minimal resources. Therefore, when designing biomarkers or screening measures, it is important to consider to what extent the test can be implemented globally with limited expense and expertise.

Limitations of This Review

This review was not intended to be a comprehensive review of each biomarker as several studies were eliminated that characterized biomarkers in different populations, established group differences, or estimated costs associated with implementation. Several existing reviews have provided in-depth discussions of one or more biomarkers or screening tools with greater details on the ease of collection, detection windows, limits of detection, costs, and feasibility of use.¹⁰¹⁻¹⁰⁴ This article aimed to focus on the predictive characteristics of biomarkers and screening tools to assess PAE and its associated impact. The search process using the selected terms may have missed relevant articles as several additional papers were found among the references in those articles identified using the initial search terms. Also, most biomarkers did not have sufficient numbers of studies for a true meta-analysis given the variation in threshold or cutoff values used to define risk and in the predictor. As a result, the range and median value of data obtained from the articles were provided. Providing uniform data-reporting formats in future studies would help with subsequent attempts to integrate these types of studies.

Future Directions

The relative importance of the predictive validity characteristics depends on the goals of the screening and on the diagnostic algorithm in which the biomarker or screening tool is being used. PPV and NPV only incorporate validity of a positive or negative test signal, respectively, and are most useful for clinicians trying to interpret a biomarker or screening tool result relative to a clinical endpoint. Accuracy provides a summary of the overall correctness of the biomarker or screening tool, but does not fully capture its errors (i.e., false positives and missed cases). In cases where the costs of these errors are high, accuracy is an inadequate indicator of success. One could argue that this is the case for PAE and its associated impact, where false positives could potentially be stigmatizing and missed cases would limit opportunities for harm reduction and intervention during early

periods of neuroplasticity. Many biomarkers and screening tools related to PAE have good specificity, but their implementation requires further evaluation of the cost-benefit ratios of use within given environments and discussions regarding the ethics of implementation relative to patient privacy and autonomy. Much progress is needed in the development of biomarkers and screening tools to improve sensitivity, which is likely to be most valued by individuals affected by PAE and those who care for them as low sensitivity results in lost opportunities for harm reduction and early intervention. AUC values provide a tool for estimating predictions that capture both sensitivity and specificity elements but may obscure relative weaknesses in one or the other. Ultimately, final decisions on clinical implementation should include input from key stakeholders who may assign different value judgments to these predictive characteristics.

Improvements in the predictive characteristics of biomarkers and screening tools would have important ramifications for surveillance methods and clinical care of individuals negatively impacted by PAE. Surveillance methods that use biomarkers or screening tools currently are limited by the low sensitivity of most available biomarkers and screening tools because a negative test result does not exclude individuals who may be negatively impacted by PAE. Surveillance studies that assume those who screen negative are unaffected and do not conduct further evaluations therefore may be underestimating true case prevalence rates. The clinical use of biomarker or screening tools also has been limited by insufficient data on predictive utility characteristics in published studies. Moreover, implementation within clinical environments often only takes place if researchers are exploring the use of the biomarker or screening tool in their studies. Improved reporting of the predictive validity characteristics of these measures are needed before consensus could be reached to support larger-scale implementation of these biomarkers and screening tools.

The field of alcohol teratogenesis initially sought to determine if PAE resulted in group differences from offspring not exposed to PAE on a variety of outcomes; however, future efforts also need to include efforts to help identify affected individuals. Predictive validity information moves beyond mean group differences and attempts to determine if a given measure's dispersion is such that a threshold, cutoff value, or rule based on an outcome or a cluster of outcomes could be used to identify those impacted by PAE. In most cases, these differing aims could be achieved within the same study, using different analyses to help with identifying better biomarkers that can improve early identification and access to rehabilitative care.

There are many promising areas where group differences have been explored but predictive characteristics have not yet been reported. One promising diagnostic tool may involve functional near-infrared spectroscopy,^{105,106} which assesses changes in oxygenation levels of brain tissue by shining near-infrared light through the scalp that is then

differentially reflected back to a sensor as different light wavelengths depending on whether or not the blood is oxygenated. Individuals with FASD show specific patterns of buildup of deoxygenated hemoglobin over time in response to prefrontal cortex activation that differ both from typically developing children and from those with other neurobehavioral impairments. Epigenetic changes, including DNA methylation, histone modifications, and other miRNAs associated with PAE, may also be effective biomarkers,^{107,108} although diagnostic analyses of these measures have rarely been reported. One promising study assessing changes in DNA methylation (i.e., the process of adding methyl groups to a DNA molecule) found that children with PAE and their mothers both had higher DNA methylation levels of proopiomelanocortin and PER2, a gene involved in regulating circadian rhythms, resulting in reduced expression of these genes. In contrast, postnatal choline supplementation, which increases the bioavailability of additional methyl groups after birth, resulted in reduced DNA methylation and increased expression levels of these stress-regulatory genes.¹⁰⁹ In addition, the health consequences of PAE are just beginning to be explored, and it may be important to determine to what extent these consequences may help identify individuals impacted by PAE.

Going beyond group differences to establish the diagnostic test validity of an outcome relative to healthy children without PAE and then relative to other children with other neurobehavioral conditions will provide the needed information to evaluate effectively whether these potential biomarkers will have clinical utility and should be further evaluated in the context of a biomarker clinical trial. This transition to a systematic process of biomarker and screening tool evaluation is needed to address the public health need of improving early identification of individuals impacted by PAE so that harm reduction and rehabilitative care efforts can be implemented.

References

1. Martin SC. *The Sage Encyclopedia of Alcohol: Social, Cultural, and Historical Perspectives*. Thousand Oaks, CA: Sage; 2015.
2. Koren G, Navioz Y. Historical perspective: The original description of fetal alcohol spectrum disorder in France, 1967. *Ther Drug Monit*. 2003;25(2):131. <https://doi.org/10.1097/00007691-200304000-00001>.
3. Lemoine P, Harousseau H, Borteyru JP, Menuet JC. Children of alcoholic parents - anomalies in 127 cases. *Arch Fr Pediatr*. 1968;25(7):830.
4. Jones KL, Smith DW. Recognition of fetal alcohol syndrome in early infancy. *Lancet*. 1973;302(7836):999-1001. [https://doi.org/10.1016/s0140-6736\(73\)91092-1](https://doi.org/10.1016/s0140-6736(73)91092-1).
5. Stratton KR, Howe CJ, Battaglia FC, Institute of Medicine; Division of Biobehavioral Sciences and Mental Disorders; Committee to Study Fetal Alcohol Syndrome, National Institute on Alcohol Abuse and Alcoholism. *Fetal Alcohol Syndrome: Diagnosis, Epidemiology, Prevention, and Treatment*. Washington, DC: National Academy Press; 1996.
6. Hoyme HE, Kalberg WO, Elliott AJ, et al. Updated clinical guidelines for diagnosing fetal alcohol spectrum disorders. *Pediatrics*. 2016;138(2):e20154256. <https://doi.org/10.1542/peds.2015-4256>.
7. Hoyme HE, May PA, Kalberg WO, et al. A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: Clarification of the 1996 Institute of Medicine criteria. *Pediatrics*. 2005;115(1):39-47. <https://doi.org/10.1542/peds.2004-0259>.
8. Chudley AE, Conry J, Cook JL, et al. Fetal alcohol spectrum disorder: Canadian guidelines for diagnosis. *CMAJ*. 2005;172(5 Suppl):S1-S21. <https://doi.org/10.1503/cmaj.1040302>.
9. Cook JL, Green CR, Lilley CM, et al. Fetal alcohol spectrum disorder: A guideline for diagnosis across the lifespan. *CMAJ*. 2016;188(3):191-197. <https://doi.org/10.1503/cmaj.141593>.
10. Astley SJ. Validation of the fetal alcohol spectrum disorder (FASD) 4-digit diagnostic code. *J Popul Ther Clin Pharmacol*. 2013;20(3):e416-e467. <https://pubmed.ncbi.nlm.nih.gov/24323701>.
11. Astley SJ, Clarren SK. Diagnosing the full spectrum of fetal alcohol-exposed individuals: Introducing the 4-Digit Diagnostic Code. *Alcohol Alcohol*. 2000;35(4):400-410. <https://doi.org/10.1093/alcalc/35.4.400>.
12. Coles CD, Gailey AR, Mulle JG, Kable JA, Lynch ME, Jones KL. A comparison among 5 methods for the clinical diagnosis of fetal alcohol spectrum disorders. *Alcohol Clin Exp Res*. 2016;40(5):1000-1009. <https://doi.org/10.1111/acer.13032>.
13. Bower C, Elliott EJ, Zimmet M, et al. Australian guide to the diagnosis of foetal alcohol spectrum disorder: A summary. *J Paediatr Child Health*. 2017;53(10):1021-1023. <https://doi.org/10.1111/jpc.13625>.
14. Landgraf MN, Nothacker M, Heinen F. Diagnosis of fetal alcohol syndrome (FAS): German guideline version 2013. *Eur J Paediatr Neurol*. 2013;17(5):437-446. <https://doi.org/10.1016/j.ejpn.2013.03.008>.
15. Bower C, Elliot EJ on behalf of the Steering Group. Report to the Australian Government Department of Health. *Australian Guide to the Diagnosis of Fetal Alcohol Spectrum Disorder (FASD)*. 2016. https://www.fasdhub.org.au/siteassets/pdfs/australian-guide-to-diagnosis-of-fasd_all-appendices.pdf.
16. Department of Health & Social Care. Report from the Guidelines Development Group to the UK Chief Medical Officers. *Alcohol Guidelines Review*. 2016. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/545739/GDG_report-Jan2016.pdf.
17. Wise J. Guidance on fetal alcohol syndrome aims to improve diagnosis and treatment in Scotland. *BMJ*. 2019;364:l396. <https://doi.org/10.1136/bmj.l396>.
18. May PA, Chambers CD, Kalberg WO, et al. Prevalence of fetal alcohol spectrum disorders in 4 US communities. *JAMA*. 2018;319(5):474-482. <https://doi.org/10.1001/jama.2017.21896>.
19. Lange S, Probst C, Gmel G, Rehm J, Burd L, Popova S. Global prevalence of fetal alcohol spectrum disorder among children and youth: A systematic review and meta-analysis. *JAMA Pediatr*. 2017;171(10):948-956. <https://doi.org/10.1001/jamapediatrics.2017.1919>.
20. Bax AC, Geurts CD, Balachova TN. Improving recognition of children affected by prenatal alcohol exposure: Detection of exposure in pediatric care. *Curr Dev Disord Rep*. 2015;2(3):165-174. <https://doi.org/10.1007/s40474-015-0057-3>.
21. Chasnoff IJ, Wells AM, King L. Misdiagnosis and missed diagnoses in foster and adopted children with prenatal alcohol exposure. *Pediatrics*. 2015;135(2):264-270. <https://doi.org/10.1542/peds.2014-2171>.
22. Strimbu K, Tavel JA. What are biomarkers? *Curr Opin HIV AIDS*. 2010;5(6):463-466. <https://doi.org/10.1097/coh.0b013e32833ed177>.

23. Fox SE, Levitt P, Nelson CA 3rd. How the timing and quality of early experiences influence the development of brain architecture. *Child Dev.* 2010;81(1):28-40. <https://doi.org/10.1111/j.1467-8624.2009.01380.x>.
24. Kable JA, O'Connor MJ, Olson HC, et al. Neurobehavioral disorder associated with prenatal alcohol exposure (ND-PAE): Proposed DSM-5 diagnosis. *Child Psychiatry Hum Dev.* 2016;47(2):335-346. <https://doi.org/10.1007/s10578-015-0566-7>.
25. Kodituwakku PW. Defining the behavioral phenotype in children with fetal alcohol spectrum disorders: A review. *Neurosci Biobehav Rev.* 2007;31(2):192-201. <https://doi.org/10.1016/j.neubiorev.2006.06.020>.
26. Kwak HS, Han JY, Choi JS, et al. Dose-response and time-response analysis of total fatty acid ethyl esters in meconium as a biomarker of prenatal alcohol exposure. *Prenat Diagn.* 2014;34(9):831-838. <https://doi.org/10.1002/pd.4374>.
27. Andreu-Fernández V, Bastons-Compta A, Navarro-Tapia E, Sailer S, Garcia-Algar O. Serum concentrations of IGF-I/IGF-II as biomarkers of alcohol damage during foetal development and diagnostic markers of foetal alcohol syndrome. *Sci Rep.* 2019;9(1):1562. <https://doi.org/10.1038/s41598-018-38041-0>.
28. Bearer CF, Santiago LM, O'Riordan MA, Buck K, Lee SC, Singer LT. Fatty acid ethyl esters: Quantitative biomarkers for maternal alcohol consumption. *J Pediatr.* 2005;146(6):824-830. <https://doi.org/10.1016/j.jpeds.2005.01.048>.
29. Gauthier TW, Mohan SS, Gross TS, Harris FL, Guidot DM, Brown LA. Placental fatty acid ethyl esters are elevated with maternal alcohol use in pregnancies complicated by prematurity. *PLoS One.* 2015;10(5):e0126552. <https://doi.org/10.1371/journal.pone.0126552>.
30. Goecke TW, Burger P, Fasching PA, et al. Meconium indicators of maternal alcohol abuse during pregnancy and association with patient characteristics. *Biomed Res Int.* 2014;2014:702848. <https://doi.org/10.1155/2014/702848>.
31. Himes SK, Dukes KA, Tripp T, et al. Clinical sensitivity and specificity of meconium fatty acid ethyl ester, ethyl glucuronide, and ethyl sulfate for detecting maternal drinking during pregnancy. *Clin Chem.* 2015;61(3):523-532. <https://doi.org/10.1373/clinchem.2014.233718>.
32. Lee HS, Kim YH, Kwak HS, Han JY, Jo SJ, Lee HK. Association of fatty acid ethyl esters in meconium of neonates with growth deficits at birth: A prospective, single-centre cohort study. *J Korean Med Sci.* 2018;33(50):e318. <https://doi.org/10.3346/jkms.2018.33.e318>.
33. Ostrea EM Jr, Hernandez JD, Bielawski DM, et al. Fatty acid ethyl esters in meconium: Are they biomarkers of fetal alcohol exposure and effect? *Alcohol Clin Exp Res.* 2006;30(7):1152-1159. <https://doi.org/10.1111/j.1530-0277.2006.00131.x>.
34. Bearer CF, Lee S, Salvator AE, et al. Ethyl linoleate in meconium: A biomarker for prenatal ethanol exposure. *Alcohol Clin Exp Res.* 1999;23(3):487-493. <https://doi.org/10.1111/j.1530-0277.1999.tb04142.x>.
35. Bakhireva LN, Leeman L, Savich RD, et al. The validity of phosphatidylethanol in dried blood spots of newborns for the identification of prenatal alcohol exposure. *Alcohol Clin Exp Res.* 2014;38(4):1078-1085. <https://doi.org/10.1111/acer.12349>.
36. Bearer CF, Jacobson JL, Jacobson SW, et al. Validation of a new biomarker of fetal exposure to alcohol. *J Pediatr.* 2003;143(4):463-469. [https://doi.org/10.1067/s0022-3476\(03\)00442-6](https://doi.org/10.1067/s0022-3476(03)00442-6).
37. Chan D, Bar-Oz B, Pellerin B, et al. Population baseline of meconium fatty acid ethyl esters among infants of nondrinking women in Jerusalem and Toronto. *Ther Drug Monit.* 2003;25(3):271-278. <https://doi.org/10.1097/00007691-200306000-00004>.
38. Bakdash A, Burger P, Goecke TW, et al. Quantification of fatty acid ethyl esters (FAEE) and ethyl glucuronide (EtG) in meconium from newborns for detection of alcohol abuse in a maternal health evaluation study. *Anal Bioanal Chem.* 2010;396(7):2469-2477. <https://doi.org/10.1007/s00216-010-3474-5>.
39. Cabarcos P, Tabernero MJ, Otero JL, et al. Quantification of fatty acid ethyl esters (FAEE) and ethyl glucuronide (EtG) in meconium for detection of alcohol abuse during pregnancy: Correlation study between both biomarkers. *J Pharm Biomed Anal.* 2014;100:74-78. <https://doi.org/10.1016/j.jpba.2014.07.023>.
40. Wurst FM, Kelso E, Weinmann W, Pragst F, Yegles M, Sundström Poromaa I. Measurement of direct ethanol metabolites suggests higher rate of alcohol use among pregnant women than found with the AUDIT—a pilot study in a population-based sample of Swedish women. *Am J Obstet Gynecol.* 2008;198(4):407.e1-407.e5. <https://doi.org/10.1016/j.ajog.2007.10.801>.
41. Gutierrez HL, Hund L, Shrestha S, et al. Ethylglucuronide in maternal hair as a biomarker of prenatal alcohol exposure. *Alcohol.* 2015;49(6):617-623. <https://doi.org/10.1016/j.alcohol.2015.06.002>.
42. Howlett H, Mackenzie S, Gray WK, et al. Assessing prevalence of alcohol consumption in early pregnancy: Self-report compared to blood biomarker analysis. *Eur J Med Genet.* 2018;61(9):531-538. <https://doi.org/10.1016/j.ejmg.2018.05.009>.
43. Niemelä O, Niemelä S, Ritvanen A, et al. Assays of gamma-glutamyl transferase and carbohydrate-deficient transferrin combination from maternal serum improve the detection of prenatal alcohol exposure. *Alcohol Clin Exp Res.* 2016;40(11):2385-2393. <https://doi.org/10.1111/acer.13207>.
44. Niemelä S, Niemelä O, Ritvanen A, et al. Fetal alcohol syndrome and maternal alcohol biomarkers in sera: A register-based case-control study. *Alcohol Clin Exp Res.* 2016;40(7):1507-1514. <https://doi.org/10.1111/acer.13101>.
45. Sarkola T, Eriksson CJ, Niemelä O, Sillanaukee P, Halmesmäki E. Mean cell volume and gamma-glutamyl transferase are superior to carbohydrate-deficient transferrin and hemoglobin-acetaldehyde adducts in the follow-up of pregnant women with alcohol abuse. *Acta Obstet Gynecol Scand.* 2000;79(5):359-366.
46. Azurmendi-Funes ML, Martínez-Villanueva M, Delgado-Marín JL, et al. An integrative screening tool of alcohol exposure during early pregnancy: Combining of the CDT biomarker with green page questionnaire. *Alcohol Alcohol.* 2019;54(6):599-608. <https://doi.org/10.1093/alcalc/agg2073>.
47. Eichler A, Grunitz J, Grimm J, et al. Did you drink alcohol during pregnancy? Inaccuracy and discontinuity of women's self-reports: On the way to establish meconium ethyl glucuronide (EtG) as a biomarker for alcohol consumption during pregnancy. *Alcohol.* 2016;54:39-44. <https://doi.org/10.1016/j.alcohol.2016.07.002>.
48. Ferraguti G, Ciolli P, Carito V, et al. Ethylglucuronide in the urine as a marker of alcohol consumption during pregnancy: Comparison with four alcohol screening questionnaires. *Toxicol Lett.* 2017;275:49-56. <https://doi.org/10.1016/j.toxlet.2017.04.016>.
49. Gomez-Roig MD, Marchei E, Sabra S, et al. Maternal hair testing to disclose self-misreporting in drinking and smoking behavior during pregnancy. *Alcohol.* 2018;67:1-6. <https://doi.org/10.1016/j.alcohol.2017.08.010>.
50. Joya X, Marchei E, Salat-Batlle J, et al. Fetal exposure to ethanol: Relationship between ethyl glucuronide in maternal hair during pregnancy and ethyl glucuronide in neonatal meconium. *Clin Chem Lab Med.* 2016;54(3):427-435. <https://doi.org/10.1515/cclm-2015-0516>.
51. Lamy S, Hennart B, Houivet E, et al. Assessment of tobacco, alcohol and cannabinoid metabolites in 645 meconium samples of newborns compared to maternal self-reports. *J Psychiatr Res.* 2017;90:86-93. <https://doi.org/10.1016/j.jpsychires.2017.02.012>.
52. May PA, Hasken JM, De Vries MM, et al. A utilitarian comparison of two alcohol use biomarkers with self-reported drinking history collected in antenatal clinics. *Reprod Toxicol.* 2018;77:25-32. <https://doi.org/10.1016/j.reprotox.2018.02.002>.
53. Maxwell S, Thompson S, Zakko F, Bracero LA. Screening for prenatal alcohol exposure and corresponding short-term neonatal outcomes. *Reprod Toxicol.* 2019;85:6-11. <https://doi.org/10.1016/j.reprotox.2019.01.009>.

87. Lucas BR, Latimer J, Fitzpatrick JP, et al. Soft neurological signs and prenatal alcohol exposure: A population-based study in remote Australia. *Dev Med Child Neurol.* 2016;58(8):861-867. <https://doi.org/10.1111/dmcn.13071>.
88. Goh PK, Doyle LR, Glass L, et al. A decision tree to identify children affected by prenatal alcohol exposure. *J Pediatr.* 2016;177:121-127. E1. <https://doi.org/10.1016/j.jpeds.2016.06.047>.
89. Mattson SN, Roesch SC, Glass L, et al. Further development of a neurobehavioral profile of fetal alcohol spectrum disorders. *Alcohol Clin Exp Res.* 2013;37(3):517-528. <https://doi.org/10.1111/j.1530-0277.2012.01952.x>.
90. Bernes GA, Courchesne-Krak NS, Hyland MT, et al. Development and validation of a postnatal risk score that identifies children with prenatal alcohol exposure. *Alcohol Clin Exp Res.* 2022;46(1):52-65. <https://doi.org/10.1111/acer.14749>.
91. Coles CD, Kalberg W, Kable JA, Tabachnick B, May PA, Chambers CD. Characterizing alcohol-related neurodevelopmental disorder: Prenatal alcohol exposure and the spectrum of outcomes. *Alcohol Clin Exp Res.* 2020;44(6):1245-1260. <https://doi.org/10.1111/acer.14325>.
92. Johnston D, Branton E, Rasmuson L, Schell S, Gross DP, Pritchard-Wiart L. Accuracy of motor assessment in the diagnosis of fetal alcohol spectrum disorder. *BMC Pediatr.* 2019;19(1):171. <https://doi.org/10.1186/s12887-019-1542-3>.
93. Mattson SN, Roesch SC, Fagerlund A, et al. Toward a neurobehavioral profile of fetal alcohol spectrum disorders. *Alcohol Clin Exp Res.* 2010;34(9):1640-1650. <https://doi.org/10.1111/j.1530-0277.2010.01250.x>.
94. Thorne JC, Coggins T. A diagnostically promising technique for tallying nominal reference errors in the narratives of school-aged children with foetal alcohol spectrum disorders (FASD). *Int J Lang Commun Disord.* 2008;43(5):570-594. <https://doi.org/10.1080/13682820701698960>.
95. Thorne JC, Coggins TE, Carmichael Olson H, Astley SJ. Exploring the utility of narrative analysis in diagnostic decision making: Picture-bound reference, elaboration, and fetal alcohol spectrum disorders. *J Speech Lang Hear Res.* 2007;50(2):459-474. [https://doi.org/10.1044/1092-4388\(2007\)032](https://doi.org/10.1044/1092-4388(2007)032).
96. Gareri J, Lynn H, Handley M, Rao C, Koren G. Prevalence of fetal ethanol exposure in a regional population-based sample by meconium analysis of fatty acid ethyl esters. *Ther Drug Monit.* 2008;30(2):239-245. <https://doi.org/10.1097/ftd.0b013e318167cfe5>.
97. Chan D, Klein J, Karaskov T, Koren G. Fetal exposure to alcohol as evidenced by fatty acid ethyl esters in meconium in the absence of maternal drinking history in pregnancy. *Ther Drug Monit.* 2004;26(5):474-481. <https://doi.org/10.1097/00007691-200410000-00003>.
98. Zelner I, Shor S, Lynn H, et al. Neonatal screening for prenatal alcohol exposure: Assessment of voluntary maternal participation in an open meconium screening program. *Alcohol.* 2012;46(3):269-276. <https://doi.org/10.1016/j.alcohol.2011.09.029>.
99. Corrigan PW, Lara JL, Shah BB, Mitchell KT, Simmes D, Jones KL. The public stigma of birth mothers of children with fetal alcohol spectrum disorders. *Alcohol Clin Exp Res.* 2017;41(6):1166-1173. <https://doi.org/10.1111/acer.13381>.
100. Alberg AJ, Park JW, Hager BW, Brock MV, Diener-West M. The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. *J Gen Intern Med.* 2004;19(5 Pt 1):460-465. <https://doi.org/10.1111/j.1525-1497.2004.30091.x>.
101. Bager H, Christensen LP, Husby S, Bjerregaard L. Biomarkers for the detection of prenatal alcohol exposure: A review. *Alcohol Clin Exp Res.* 2017;41(2):251-261. <https://doi.org/10.1111/acer.13309>.
102. Lim YH, Watkins RE, Jones H, Kippin NR, Finlay-Jones A. Fetal alcohol spectrum disorders screening tools: A systematic review. *Res Dev Disabil.* 2022;122:104168. <https://doi.org/10.1016/j.ridd.2021.104168>.
103. Bortolotti F, Tagliaro F. Biomarkers for the identification of alcohol use/abuse: A critical review. *Forensic Sci Rev.* 2011;23(2):55-72.
104. Burd L, Hofer R. Biomarkers for detection of prenatal alcohol exposure: A critical review of fatty acid ethyl esters in meconium. *Birth Defects Res A Clin Mol Teratol.* 2008;82(7):487-493. <https://doi.org/10.1002/bdra.20464>.
105. Barrett CE, Kable JA, Madsen TE, Hsu CC, Coles CD. The use of functional near-infrared spectroscopy to differentiate alcohol-related neurodevelopmental impairment. *Dev Neuropsychol.* 2019;44(2):203-219. <https://doi.org/10.1080/87565641.2019.1567734>.
106. Kable JA, Coles CD, CIFASD. Prefrontal cortical responses in children with prenatal alcohol-related neurodevelopmental impairment: A functional near-infrared spectroscopy study. *Clin Neurophysiol.* 2017;128(11):2099-2109. <https://doi.org/10.1016/j.clinph.2017.08.009>.
107. Lussier AA, Bodnar TS, Weinberg J. Intersection of epigenetic and immune alterations: Implications for fetal alcohol spectrum disorder and mental health. *Front Neurosci.* 2021;15:788630. <https://doi.org/10.3389/fnins.2021.788630>.
108. Resendiz M, Chen Y, Oztürk NC, Zhou FC. Epigenetic medicine and fetal alcohol spectrum disorders. *Epigenomics.* 2013;5(1):73-86. <https://doi.org/10.2217/epi.12.80>.
109. Sarkar DK, Gangisetty O, Wozniak JR, et al. Persistent changes in stress-regulatory genes in pregnant women or children exposed prenatally to alcohol. *Alcohol Clin Exp Res.* 2019;43(9):1887-1897. <https://doi.org/10.1111/acer.14148>.

Appendix 1. Predictive Characteristics* of Biomarkers for Prenatal Alcohol Exposure Status

Author	Subject Description	PAE/FASD Sample Description	Control Sample Description	Biomarker	Source of Biomarker	Outcome	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC From ROC
Azurmendi-Funes et al., 2019 ⁴⁶	Mothers at a public maternity hospital in Spain	20 High risk due to PAE (20 g alcohol or ≥ 3 binges of ≥ 40 g alcohol)	71 Non-high-risk PAE	%CDT (cutoff ≥ 95%) AST ALT GGT MCV	Maternal blood	Self-report of PAE	25.0	93.0	81.5			.75/.71 .47 .54 .42 .38
Bakdash et al., 2010 ³⁸	Meconium samples	116 FAEE (≥ 100 ng/g) 28 PAE	480 FAEE (< 100 ng/g) 32 No PAE	EtG ≥ 500 ng/g	Meconium	FAEE > 100 ng/g	78.6	98.7	82.5	98.4	97.3	
Bakhireva et al., 2014 ³⁵	Women from a speciality treatment clinic			PEth > 8 ng/mL delivery GGT > 40 U/L mid-gestation GGT > 40 U/L delivery %CDT > 2.0% mid-gestation %CDT > 2.0% delivery EtG ≥ 25 ng/mL mid-gestation EtG ≥ 25 ng/mL delivery EtS ≥ 7 mid-gestation EtS ≥ 7 delivery PEth (> 8 ng/mL) FAEE (sum of 9) All mid-gestation	Maternal blood Maternal urine Infant blood spots Infant meconium	Self-report of PAE	22.2 14.8 10.7 3.7 10.7 14.8 7.4 14.8 7.4 32.1 28.6 32.1	100 100 86.2 100 96.7 96.9 96.9 100 96.9 100 81.3 96.9				

Appendix 2. Predictive Characteristics* of Biomarkers for Fetal Alcohol Spectrum Disorders and Associated Symptoms

Author	Sample Description	PAE/FASD	Control Sample	Biomarker	Source of Biomarker	Outcome	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC from ROC
Andreu et al., 2019 ²⁷	55 native Spanish children (ages 8–12) from the Meconium Project and 98 children adopted from Eastern European communities	55 native Spanish children with FASD; 33 FAE > 2 nmol/g; of 98 children from Eastern Europe: 31 with FASD, 42 partial FAS, 6 ARBD, and 5 ARND	31 FAE < 2 nmol/g	Serum IGF-II (below 5th percentile)	Infant blood	FASD diagnosis	12.7	100	100	14.8	24.2	
Balaraman et al., 2016 ⁵⁷	Ukrainian pregnant mothers at prenatal care visits	22 Heavy alcohol-affected	23 No PAE	Serum IGF-II (below 5th percentile) Selected mRNAs mid to late pregnancy Change in mRNAs	Infant blood	FASD diagnosis	39.0	96.8	98.8	19.4	46.6	
Lee et al., 2018 ³²	Meconium samples from infants born in hospitals—Uijeongbu St. Mary's Hospital, Catholic University of Korea	5 with HC and length ≤ 10th percentile	FAEE < 0.5 nmol/g FAEE < 2 nmol/g FAEE < 0.5 nmol/g FAEE < 2 nmol/g	FAEE 0.5 nmol/g FAEE 2 nmol/g FAEE 0.5 nmol/g FAEE 2 nmol/g	Meconium Meconium Meconium Meconium	Growth delay—HC, length < 10th percentile	40.0 20.0 50.0 25.0	92.8 98.0 92.8 98.0			68.0 68.0 79.0 79.0	

Appendix 3. Predictive Characteristics* of Other Screening Tools for Fetal Alcohol Spectrum Disorders and Its Associated Symptoms

Author	PAE/FASD Sample	Comparison Sample	Predictor	Outcome	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC from ROC
Aring et al., 2021 ⁶⁶	37 Children with FASD	65 Healthy children, 33 ADHD, 57 Moderate to late prematurity 65 Healthy children only 16 Silver-Russell syndrome 65 Healthy children All comparison groups 33 ADHD 57 Moderate to late prematurity 65 Healthy children only	FASD Eye Code ≥ 10	FASD	43 43 43 57	94 100 88 98				0.78 0.87 0.6 0.92 0.76 0.66 0.75 0.87
Astley & Clarren, 1996 ⁵⁸	42 FAS; 21 development sample and 21 validation sample	84 without FAS (including 4 with other genetic conditions) placed into 2 groups; 42 per group for development and then validation	Facial features: 2D continuous measurements philtrum/lip Likert scale rating philtrum & lip Likert scale philtrum/lip continuous	Gestalt FAS	100 100 100	93 100 100				
Astley et al., 2002 ⁵⁹	Sampled 600 children in foster care screened	Facial analysis software	Facial features: 2D clinical FASD diagnosis	Screened + Gestalt FAS	100	99.8	85.7	100	99.8	
Moore et al., 2001 ⁶⁰	41 FAS & 59 pFAS	31 Controls	Facial features: 6 2D craniofacial measurements 2 Craniofacial measurements 5 Craniofacial measurements	Clinical FAS/pFAS	98 100 86	90 100 94			96 100 88	
Widder et al., 2021 ⁶¹	22 FASD	31 Controls; 15 ADHD; 20 AUD/ODD; 18 depression	Facial features: 2D German BSI-FASD Facial analysis software German BSI-FASD adapted scoring	Clinical FASD	77 67 86	70-100 44-79 70-100				

