Am. J. Hum. Genet. 46:587–590, 1990

Maternal Serum Alpha-Fetoprotein (MSAFP) Patient-Specific Risk Reporting: Its Use and Misuse

James N. Macri,* Ramana V. Kasturi,* David A. Krantz,* Edward J. Cook,* and John W. Larsen†

*Research Division, NTD Laboratories, Inc., Carle Place, NY; and †Department of Obstetrics, Gynecology, and Genetics, George Washington University Medical Center, Washington, DC

Summary

Fundamental to maternal serum alpha-fetoprotein screening is the clinical utility of the laboratory report. It follows that the scientific form of expression in that report is vital. Professional societies concur that patient-specific risk reporting is the preferred form. However, some intermediate steps being taken to calculate patient-specific risks are invalid because of the erroneous assumption that multiples of the median (MoMs) represent an interlaboratory common currency. The numerous methods by which MoMs may be calculated belie the foregoing assumption.

Introduction

Some presymptomatic screening tests (unlike diagnostic tests) do not establish the presence of a condition but, instead, allow the physician and genetic counselor to revise the probability that the condition exists (Macri et al. 1987*a*). Reporting such screening results as patient-specific risks facilitates decision making by physicians and patients (Pauker and Kassirer 1987*a*, 1987*b*).

The evaluation of maternal serum alpha-fetoprotein (MSAFP) during the second trimester of pregnancy, for the early identification of open neural tube defects, is a population-based screening procedure taking place in wanted pregnancies. It leads, in some cases, to invasive diagnostic procedures which have documented risks to maternal and fetal well-being. Laboratories, therefore, should report MSAFP screening results in probabilistic terms, a form of reporting not employed in diagnostic testing.

Recognizing the need for probabilistic reporting in the clinical interpretation of MSAFP determinations, the Centers for Disease Control, in 1984, published a framework by which laboratories might adopt this practice (Adams et al. 1984). Evaluation and implementation of patient-specific risk reporting began in our laboratories in 1985, followed by our report (Macri 1986) on this aspect of MSAFP screening. Subsequent to these publications, leading professional societies (American College of Obstetricians and Gynecologists [1986]; American Academy of Pediatrics, Committee on Genetics [1987]; and American Society of Human Genetics [1987]) concluded that probabilistic reporting represents the preferred mode of conveying MSAFP screening information to both the clinician and patient.

The intervening years have seen a gradual adoption of patient-specific risk reporting in MSAFP screening. In some instances, however, the transition from reporting quantitative levels of MSAFP to reporting patient-specific risks has not been made on the basis of scientifically sound methods. The difficulty lies in the assumption that an intermediate statistical expression (multiple of the median [MoM]) represents an interlaboratory common currency (Knight et al. 1988). To test the validity of this assumption we undertook a review of the extent of variation in MoMs attributable to the methods employed in calculating them.

Interalgorithm Variation of MoMs

Laboratories that use MoMs must decide on an algorithm with which to compute them. Such a decision

Received September 22, 1989.

Address for correspondence and reprints: James N. Macri, Ph.D., Research Division, NTD Laboratories, Inc., 383 Old Country Road, Carle Place, NY 11514.

0002-9297/90/4603-0023$02.00

can include (1) how to establish normative medians at each gestational age, (2) whether the laboratory should adjust for maternal weight, and, if so, (3) which weight-adjustment approach should be used (Wald et al. 1981; Palomaki et al. 1985; Macri et al. 1986; Kazazian et al. 1987; Macri et al. 1987*d*). In addition to these decisions, laboratories are faced with alternatives in adjusting for ethnicity, insulin dependency, and multiple gestation.

Three common methods exist for the establishment of medians. The first, used by many low-volume laboratories, is to set normative medians equal to the package-insert medians which manufacturers supply with their kits. A second alternative measures samples at each gestational week and calculates medians from this empirical data (observed medians). A third approach, recommended by the first United Kingdom collaborative study on alpha-fetoprotein in relation to neural tube defects (U.K. Collaborative Study 1977), calls for the computation of a regression of the logarithms of the observed medians vs. gestational age, weighted by the number of samples at each gestational week. The medians are then derived by solving the equation of the regression line at each gestational week.

The three methods for establishing medians along with adjustment for maternal weight result in twelve different algorithms to determine MoMs from quantitative data. To illustrate this point we have used published data generated for the implementation of MSAFP screening (Christensen et al. 1986). Table 1 shows how a laboratory could report any of 12 different MoM values for each of three hypothetical samples, depending on which of the 12 algorithms it chose to implement. The range of possible MoMs for each of the three samples is substantial (0.43–0.64, 1.70–2.55, and 2.13–3.19).

If decisions were made on whether to adjust for ethnicity and insulin dependency, the number of possible algorithms would increase geometrically. For example, if the decision to use a uniform adjustment factor for race (Johnson 1985, pp. 183–195), to use black medians (Macri et al. 1987*b*), or not to adjust for race is considered, the number of algorithms triples. In addition, if the decision to use a uniform MSAFP adjustment for insulin dependency (Wald et al. 1979), a diabetic gestational age adjustment (Milunsky et al. 1982), or no adjustment for insulin-dependent patients is considered, the number of algorithms triples again. As a result, samples from black gravid women who are insulin dependent could be evaluated in 108 different ways (i.e., 12 × 3 × 3). Adjustment for additional factors would likely lead to wider ranges in MoM values for the same sample.

The assumption that all MoMs are equal has led to (1) the development of laboratory proficiency testing based on MoMs and (2) the adoption of tables and

Table 1

Inter-algorithm Differences in MoMs for an 18-wk Gravida

Medians[a]	Weight Adjustment Method[b]	MoMs at Three Quantitative Levels		
		18.80 IU/ml	75.20 IU/ml	94 IU/ml
Observed	No adjustment	.50	2.0	2.50
Observed	Palomaki	.46	1.86	2.32
Observed	Wald	.45	1.78	2.23
Observed	Kazazian	.43	1.70	2.13
Regressed	No adjustment	.56	2.25	2.81
Regressed	Palomaki	.52	2.09	2.61
Regressed	Wald	.50	2.01	2.51
Regressed	Kazazian	.48	1.92	2.39
Package insert	No adjustment	.64	2.55	3.19
Package insert	Palomaki	.60	2.37	2.97
Package insert	Wald	.57	2.27	2.84
Package insert	Kazazian	.55	2.17	2.72
Range		0.43–0.64	1.70–2.55	2.13–3.19

[a] Observed and package-insert medians are from published data (Christensen et al. 1986). Regressed medians are derived from observed medians at 14–19 wk.

[b] Data based on a maternal weight of 115 lbs. MoM differences will be observed at any maternal weight.

nomograms which purport to convert the MoM of any laboratory into a patient-specific risk.

The Use of MoMs in Interlaboratory Comparison of Data

Historically MSAFP proficiency testing programs have provided us with insight into AFP MoM variability. For example, results of the 1978 Centers for Disease Control Experimental AFP Proficiency Test Program (Taylor et al. 1983) demonstrated that the use of MoMs to express AFP values "resulted in a 28% increase in variance . . . contrary to what was expected." Additionally, data from an external quality-assessment scheme in the United States demonstrated a similar increase in variance when MSAFP results were expressed in MoMs. In the program of the external quality-assessment scheme (Knight et al. 1985), laboratory MoM results varied to such a degree that this proficiency-testing program accepted values on a single proficiency-test sample which varied from 1.96 to 4.80 MoMs.

Since participating laboratories are likely to use different AFP reagent kits and also use dissimilar algorithms to compute MoMs, the MoMs produced by them (each individual laboratory) will generate disparate statistical distributions. As a result, proficiency-testing programs will be forced to choose between unfairly rejecting accurately measured samples and increasing acceptability ranges to such an extent that some laboratories are deemed qualified despite poor performance.

Use of Nomograms

Recognizing the importance of patient-specific risk reporting, some laboratories have begun to use risk tables and nomograms established by other laboratories. These devices purport to convert any laboratory's MoM findings into patient-specific risks. This procedure assumes that MoMs generated by the laboratory which created the risk table are sufficiently similar to those of any other laboratory using the risk table that an accurate risk will be produced. Data presented herein pointedly contradict that assumption.

Table 1 shows that a laboratory could determine that a serum sample measuring 18.80 IU/ml had an MoM value ranging anywhere from 0.43 to 0.64 MoM. As a result, that laboratory could report a Down syndrome risk using Palomaki's risk table (Palomaki and Haddow 1987) ranging from 1/211 to 1/466 for a 30-year-old gravida. This wide range is clinically problematic, given the current practice of offering amniocentesis when a patient's risk equals or exceeds that of a 35-year-old (1/365 at term). Similarly, if we were to use another published risk table (Fourth report of the U.K. collaborative study 1982), the risk of neural tube defects associated with a 2.55 MoM is more than five times as great as the risk associated with an MoM of 1.70 (table 1, col. 2). The variability in MoMs caused by the use of different algorithms will result in misleading indications of patient-specific risks if such risk tables and nomograms are employed.

Discussion

In a recent study, Hook (1988*a*, 1988*b*) illustrates that the relationship between MoMs and corresponding risks for Down syndrome vary to an unacceptable degree in the studies published thus far. He states, "For example, Table 1 of DiMaio et al. implies that a value for maternal serum alpha-fetoprotein of 1.3 MoM in a 39-year-old woman predicts a risk of Down syndrome comparable to that of an average 35-year-old. But other papers suggest values of 1.1 MoM, 1.5 MoM, or 1.9 MoM for such a risk. Depending on whose values are used, the proportion of 39-year-olds at or below such a risk is 7 percent (1.9 MoM), 18 percent (1.5 MoM), 27 percent (1.3 MoM), or 39 percent (1.1 MoM)."

Evans et al. (1988), evaluating 14 Detroit laboratories, report, "The wide variation found in reported values, medians, and use of correction factors makes interpretation of results difficult and inaccurate."

Data we have analyzed demonstrate that MoMs do not represent a common currency among laboratories since there exist numerous methods for computing them. This argues against the reporting of MSAFP results in quantitative terms and reinforces the point that prenatal MSAFP screening has not found a scientifically or clinically safe harbor in simulations of correctly generated patient-specific risks. In the field of prenatal MSAFP screening there are great advantages to the independent development of normative data bases, rigorous quality-control and quality-assurance procedures, assaying of large numbers of samples (Macri et al. 1987*c*), and adherence to a protocol of successive screening and diagnostic procedures initiated by reports of screening in patient-specific risks.

References

American College of Obstetricians and Gynecologists (1986) Prenatal detection of neural tube defects. Technical bulle-

tin no. 99, American College of Obstetricians and Gynecologists, Washington, DC

Adams MJ, Windham GC, James LM, Greenberg F, Clayton-Hopkins JA, Reimer CB, Oakley GP (1984) Clinical interpretation of maternal serum alpha-fetoprotein concentrations. Am J Obstet Gynecol 148:241–254

American Academy of Pediatrics, Committee on Genetics (1987) Alpha-fetoprotein screening. Pediatrics 80:444–445

American Society of Human Genetics (1987) American Society of Human Genetics policy statement for maternal serum alpha-fetoprotein screening programs and quality control for laboratories performing maternal serum and amniotic fluid alpha-fetoprotein assays. Am J Hum Genet 40:75–82

Christensen RL, Rea MR, Kessler G, Crane JO, Valdes R (1986) Implementation of a screening program for diagnosing open neural tube defects: selection, evaluation, and utilization of alpha-fetoprotein methodology. Clin Chem 32:1812–1817

Evans MI, Belsky R, Greb A, Dvorin E, Drugan A (1988) Wide variation in maternal serum alpha-fetoprotein reports in one metropolitan area: concerns for the quality of prenatal testing. Obstet Gynecol 72:342–345

Fourth report: U.K. Collaborative Study on Alpha-Fetoprotein in Relation to Neural Tube Defects (1982) Estimating an individual's risk of having a fetus with open spina bifida and the value of repeat alpha-fetoprotein testing. J Epidemiol Community Health 36:87–95

Hook EB (1988*a*) Screening for Down syndrome. N Engl J Med 318:928

Hook EB (1988*b*) Variability in predicted rates of Down syndrome associated with elevated maternal serum alpha-fetoprotein levels in older women. Am J Hum Genet 43: 160–164

Johnson AM (1985) Racial differences in maternal serum alpha-fetoprotein screening. In: Mizejewski GJ, Porter IH (eds) Alpha-fetoprotein and congenital disorders. Academic Press, New York, pp. 183–196

Kazazian LC, Niebyl JR, Joseph JM (1987) Maternal serum alpha-fetoprotein screening: effects of weight adjustment. Am J Obstet Gynecol 157:782

Knight GJ, Haddow JE, Palomaki G (1985) An external laboratory quality control program to help ensure reliability for maternal serum alpha-fetoprotein screening. Foundation for Blood Research, Scarborough, ME

Knight GJ, Palomaki GE, Haddow JE (1988) Quality assurance in MSAFP screening. Contemp Ob/Gyn 31:16–24

Macri JN (1986) Critical issues in prenatal maternal serum alpha-fetoprotein screening for genetic anomalies. Am J Obstet Gynecol 155:240–246

Macri JN, Hu MG, Krantz DA (1987*a*) Decision analysis. N Engl J Med 317:387–388

Macri JN, Kasturi RV, Hu MG, Koch KE, Krantz DA, Douros TJ, Sajda P, et al (1987*b*) Maternal serum alpha-fetoprotein screening. III. Pitfalls in evaluating black gravid women. Am J Obstet Gynecol 157:820–822

Macri JN, Kasturi RV, Krantz DA, Hu MG (1987*c*) Maternal serum alpha-fetoprotein screening. II. Pitfalls in low volume decentralized laboratory performance. Am J Obstet Gynecol 156:533–535

Macri JN, Kasturi RV, Krantz DA, Koch KE (1986) Maternal serum alpha-fetoprotein screening, maternal weight, and detection efficiency. Am J Obstet Gynecol 155:758–760

Macri JN, Kasturi RV, Krantz DA, Koch KE, Hu MG (1987*d*) Maternal serum alpha-fetoprotein screening: effects of weight adjustment. Am J Obstet Gynecol 157:783–784

Milunsky A, Alpert E, Kitzmiller JL, Younger MD, Neff RK (1982) Prenatal diagnosis of neural tube defects. VIII. The importance of serum alpha-fetoprotein screening in diabetic pregnant women. Am J Obstet Gynecol 142:1030–1032

Palomaki GE, Haddow JE (1987) Maternal serum alpha-fetoprotein, age, and Down syndrome risk. Am J Obstet Gynecol 156:460–463

Palomaki GE, Knight GJ, Kloza EM, Haddow JE (1985) Maternal weight adjustment and low serum alpha-fetoprotein values. Lancet 1:468

Pauker SG, Kassirer JP (1987*a*) Decision analysis. N Engl J Med 316:250–258

Pauker SG, Kassirer JP (1987*b*) Decision analysis. N Engl J Med 317:387–388

Taylor RN, Przybyszewski VA, Gray E (1983) Results of the Centers for Disease Control experimental proficiency testing survey for serum alpha-fetoprotein. J Clin Microbiol 17:100–105

U.K. Collaborative Study on Alpha-Fetoprotein in Relation to Neural Tube Defects (1977) Maternal serum alpha-fetoprotein measurement in antenatal screening for anencephaly and spina bifida in early pregnancy. Lancet 1: 1323–1332

Wald NJ, Cuckle H, Boreham J, Stirrat GM, Turnbull AC (1979) Maternal serum alpha-fetoprotein and diabetes mellitus. Br J Obstet Gynaecol 86:101–105

Wald N, Cuckle H, Boreham J, Terzian E (1981) The effect of maternal weight on maternal serum alpha-fetoprotein levels. Br J Obstet Gynaecol 88:1094–1096