

J. Perinat. Med.  
14 (1986) 101

## **Accuracy of an average ultrasonic laboratory in measurements of fetal biparietal diameter, head circumference and abdominal circumference**

**Umberto Nicolini, Enrico Ferrazzi, Renata Molla, Elisabetta Massa, Gabriele Cicognani, Marina Santarone, Maria Bellotti, and Giorgio Pardi**

1st Department of Obstetrics and Gynecology, University of Milan (Director Prof. G. B. Candiani), Italy

### **1 Introduction**

Ultrasonic evaluation of gestational age and fetal growth relies on measurements of one or more anatomical landmarks which are compared to standard growth curves and/or to previous measurements. The correct utilization of ultrasonic data implies full knowledge of the limits in accuracy of the measurements.

The whole question of ultrasonic accuracy has been assessed in a number of studies. Most of these have concentrated on the commonest parameter, i. e. biparietal diameter (BPD).

The earliest papers, published in the late sixties, dealt with the error due to the ultrasonic method itself and compared BPD measured ultrasonically before delivery and at birth [3, 4]. It was concluded that if the equipment is correctly adjusted, this source of inaccuracy can be assumed to be fairly constant and unimportant for practical purposes.

It was noted in several subsequent studies that accuracy is related to the reproducibility of ultrasonic measurement which largely depends on observer error, the technical specification of the equipment and the actual subject being measured. Different studies reported varying standard deviations (SD) in a series of measurements of the same BPD, according to the degree

of experience of the operators, the time interval between measurements and the type of ultrasonic equipment used [2, 9, 11, 15, 16]. When different operators were compared, a statistically significant difference was noted in nearly all cases [10, 14].

Few papers have been published regarding other anatomical landmarks such as head circumference and abdominal circumference which are commonly used in evaluating gestational age and fetal growth, but similar conclusions can be drawn [8, 14, 18, 19].

In most ultrasonic laboratories, different operators use different equipment to perform the same procedures and produce results which are commonly assumed to be homogeneous. Image processing and display, ease of alignment to predetermined scan planes and measurements facilities vary greatly from one machine to another. Moreover, the relative experience, ability, care and possible fatigue of the operator all determine the way in which predetermined fetal sections are obtained and the varying images thus produced are measured. The reliability of any ultrasound laboratory will be influenced by all these factors; and, for each parameter the reliability is probably somewhat different from that reported.

## 2 Material and methods

Twenty patients, selected from those admitted for various conditions (i. e. diabetes, toxemia, etc.) to the antenatal ward of the First Department of Obstetrics and Gynecology of the University of Milan, consented to take part in the investigation. Gestational age ranged between the 14th and 40th week (8 were less than the 28th week). Each patient was scanned by six operators on three different occasions during a single day (early morning, afternoon and evening). Three of the operators used a compound scanner (C.S.) and three a real time scanner (R.T.). Their experience varied but was not less than three months at the beginning of the study; except for one operator who used the real time scanner and only had two weeks training. On each occasion each operator had to determine biparietal diameter (BPD), head circumference (HC) and abdominal circumference (AC).

Static measurements of BPD were obtained by using the A-scan, B-scan technique [3]. Multi-directional calipers were positioned directly on the outer and inner tables on the real time video display. Caliper velocity was 1600 m/sec with both machines. HC and AC were obtained according to the CAMPBELL and THOMS [6] and CAMPBELL and WILKIN [7] techniques and measured on polaroid film with a map measurer. System velocity was 1540 m/sec with both the compound and real time scanner. A 3/5 or 4/5 scale was generally used with the compound scanner while the image size was fixed in the case of real time equipment. Measurements on polaroid film were corrected for the life size/photo ratio. All measurements were obtained "blind". "Wild" measurements were not discarded. There was no time limit, but scans seldom lasted more than twenty minutes.

Circumferences were measured immediately after the scan by the operator and a second coded copy was retained. The whole set of photographs was subsequently measured by one of the operators. Variance analysis was carried out for each parameter from the pool of measurements obtained (3 scans by 6 operators on 20 patients). Any missing data in a set

of measurements was replaced by its corresponding mean, leaving degrees of freedom unchanged. The following results are given for each parameter:

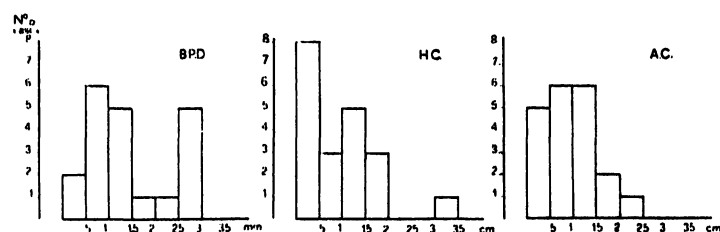
1. differences among operators using C.S. and among those using R.T.
2. mean SD of measurements obtained using C.S. (all three operators together), using R.T. (likewise) and with both machines (all six operators together): a) for all 20 patients, b) for each patient.
3. differences and interaction between measurements obtained using C.S. (all three operators together) and using R.T. (likewise).
4. variance analysis and Student's t-test were performed to evaluate any statistical differences between 720 measurements of the retained photographs of head and abdomen circumferences and the corresponding measurements carried out immediately after the procedure.

## 3 Results

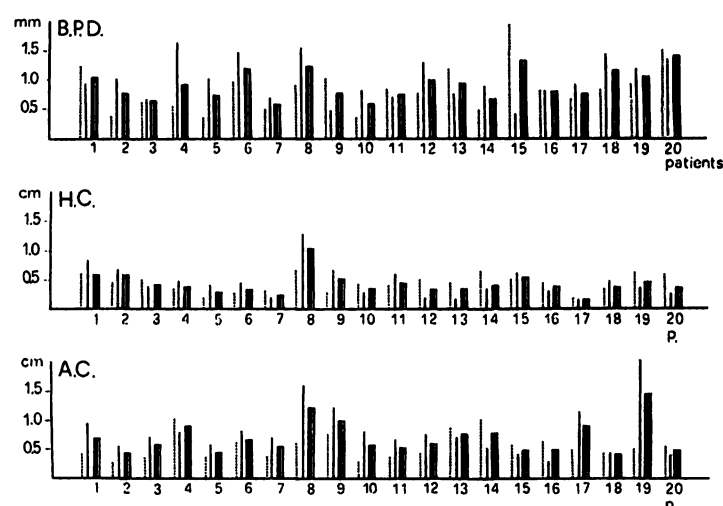
Three hundred forty eight measurements of BPD, 337 of HC and 346 of AC were obtained; the missing measurements due to technical difficulties were thus 3.3%, 6.4% and 3.9% respectively. In only two cases did two different operators using the real time scanner not obtain a complete set of measurements of one parameter (in one case HC, in the other AC).

Analysis of variance among the operators using the compound scanner revealed significant differences ( $p < 0.05$ ) with BPD but not significant with HC or AC; among the operators using the real-time scanner, differences were significant with all parameters (BPD:  $p < 0.05$ ; HC:  $p < 0.001$ ; AC:  $p < 0.001$ ). The mean values of measurements obtained for each patient by three operators using the real-time scanner differed less than 2.5 mm in 15 cases with BPD, less than 1.5 cm in 16 cases with HC and in 17 cases with AC (figure 1).

Mean SDs of measurements obtained by operators using C.S. were as follows: BPD 0.94 mm, HC 0.43 cm and AC 0.55 cm, and using R.T.:



**Figure 1.** Maximum differences between mean values of measurements obtained by 3 operators using real time scanner on 20 patients.



**Figure 2.** Mean S.D. obtained in each patient by 3 operators using the compound scanner (.....), by 3 operators using the real time scanner (—), and the 6 operators together (—).

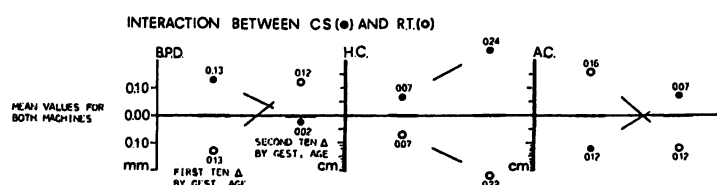
BPD 1.01 mm, HC 0.53 cm and AC 0.83 cm. The overall mean SDs (C.S. and R.T. together) were: BPD 0.97 mm, HC 0.49 cm and AC 0.77 cm.

The mean SDs of measurements obtained for each of the 20 patients on C.S., R.T. and on both machines together are set out progressively in figure 2 by gestational age. No evident relationships between the absolute value of mean SDs and gestational age can be seen ( $F$  of the straightline is not significant). The mean SDs of measurements obtained for each patient on both machines, in no case exceeded 1.5 mm with BPD and 1.0 cm with HC and were higher than 1.0 cm in only three cases with AC.

Mean values of BPD, HC and AC obtained for each patient on the C.S. and on the R.T. are

**Table I.** Mean values of B.P.D., H.C. and A.C. obtained for each patient on the compound scanner and the real time.

No.	B.P.D.		H.C.		A.C.	
	C.S.	R.T.	C.S.	R.T.	C.S.	R.T.
1	29.3	30.2	11.2	10.8	9.2	9.8
2	35.7	34.3	12.6	12.5	10.0	10.6
3	35.8	36.9	13.0	12.6	10.9	10.8
4	35.9	35.2	12.6	12.6	11.2	11.3
5	45.2	43.9	16.5	15.6	14.1	13.1
6	47.6	47.7	18.1	18.4	15.4	15.3
7	52.1	51.9	18.6	19.4	15.6	16.1
8	61.2	61.0	21.8	22.0	20.4	21.5
9	64.5	63.3	23.6	23.1	21.9	20.9
10	74.5	74.0	26.3	25.7	25.9	25.7
11	76.3	77.7	26.7	26.6	24.2	24.9
12	82.5	81.6	28.8	27.9	25.8	25.2
13	84.8	87.0	30.0	29.6	30.6	32.0
14	90.4	89.8	30.9	30.8	28.2	28.5
15	90.5	91.6	31.5	30.4	30.1	29.9
16	92.2	92.1	32.5	32.6	30.6	30.3
17	94.7	95.6	33.4	34.0	33.2	33.3
18	97.2	95.1	34.7	33.5	34.3	33.4
19	98.1	97.3	34.2	33.3	32.8	31.2
20	99.3	99.6	35.5	34.5	35.9	34.8
	p	NS	p < 0.001		p	NS



**Figure 3.** Figures on the left represent the mean of the first ten (by gestational age) differences between mean values for one machine (open circle R.T., closed circle C.S.) and mean values for both machines. The figures on the right side represent the same mean for the second differences.

reported in table I. Differences between C.S. and R.T. were not significant with BPD and AC, but were significant with HC ( $p < 0.001$ ). The interaction between both machines with all three parameters was significant at the same level of  $p < 0.001$ . The interaction between C.S. and R.T. for each parameter is shown in figure 3.

Mean values of triplets of measurements obtained immediately after the procedure by each

operator and mean values on retained photographs were significantly different in 9 out of 60 cases with HC and in 5 out of 60 with AC, using C.S.; using R.T. they were different in 8 out of 60 cases with HC and 12 out of 60 with AC. While there were no significant differences with HC or AC among operators using C.S. as between measurements obtained from line and from retained photographs, significant differences occurred among operators using R.T. (HC:  $p < 0.001$ , AC:  $p < 0.05$ ).

#### 4 Discussion

When one deals with data obtained in an ultrasound laboratory where different people are using different machines, it should be known whether there are inconsistencies among the operators and whether the routine scanners and more sophisticated equipment they use yield homogeneous results. Indeed, the measurements produced may be significantly different or vary in their degree of reproducibility. Finally it should be clarified whether gestational age i.e. the dimensions and mobility of the fetus is an important variable in the accuracy of the measurements. This latter factor is not unimportant. For example, the method of growth adjusted sonographic age (G. A. S. A.) proposed by SABBAGHA et al. [17] assumes that the accuracy of BPD remains unchanged during the whole course of pregnancy. One report gives different SDs in series of measurements of the same BPD according to the absolute value of BPD [15].

In the present study none of the three parameters showed any relationships between reproducibility and gestational age, and this applied to measurements obtained with both C.S. and R.T. (figure 2). These results are in agreement, at least as regards BPD, with other published papers [12, 14] and suggest that differences in confidence limits of standard growth curves at different gestational ages mainly reflect biological variability rather than the degree of accuracy.

SDs of BPD have been reported to vary between 0.25 mm and 2.54 mm [4, 11]. The mean

SD of measurements obtained by all six operators on different occasions with two machines in our laboratory is comparable with data in the literature; moreover, the lowest values previously reported, were obtained by single operators on the same occasion.

A SD for BPD of 0.97 mm gives a degree of accuracy which can be estimated by the following example: a BPD of 50.2 mm, which when referred to CAMPBELL's growth curve [5] corresponds to the mean for 20 weeks, might actually be 52.1 mm ( $50.2 \text{ mm} + 2 \text{ SD}$ ) or 48.3 mm ( $50.2 \text{ mm} - 2 \text{ SD}$ ) with 95% of confidence limits. The higher value may actually fit with the 5th centile for the 21.4 weeks while the lower one may fit with the 95th centile for the 18.3 weeks. In this way a measurement of BPD which gives an estimated gestational age of 20 weeks will only record approximate gestational age within plus or minus 11 days.

Concerning HC and AC, mean SD was slightly better with C.S. than with R.T. while overall SD (0.49 cm for HC and 0.77 cm for AC) is similar to the few data reported by others. Such a limit to accuracy of measurement of AC cannot rule out that a fetus estimated to have reached the 25th centile at the 30th week (26.3 cm referred to standard growth curves of SABBAGHA et al. [18]) may be, at the opposite extremes of possibility at the 50th or 5th centile.

SD, which represents the degree of reproducibility of measurements, reflects what limitations there are both in determining gestational age and evaluating fetal growth. On the basis of these findings, ultrasonic prediction of expected date of delivery (EDD) given by a routine laboratory cannot provide the same degree of accuracy as that reported by CAMPBELL and NEWMAN who have estimated gestational age within plus or minus 7 days by a single determination of BPD before 24 weeks [5]. This relatively low degree of precision is by far better than that provided by certain dates, which has been reported to vary between plus or minus 29 days [1].

The limits to reproducibility in our laboratory could also explain the higher failure rate of

routine screening of fetal growth impairment, as compared with prospective studies. In our laboratory, the percentage of the false negatives in detecting small for gestational age fetuses, from the 25th to the 40th week of gestation, is in fact as high as 44.1% [13].

In view of the influence of the two possible sources of errors described, i. e., performance and type of equipment used, our results suggest that experience of operators is less important in C.S. than in R.T. Other authors held a different opinion at least as regards BPD. As a matter of fact, previously reported SDs of measurements of BPD obtained with R.T. show a narrower range of variability than those obtained with C.S. In the present study measurements of BPD were indeed statistically different at the same level ( $p < 0.05$ ) among the operators using both C.S. and R.T., but differences among the operators using C.S. were not statistically significant with HC and AC, while they were significant, even though probably unimportant for clinical purposes, in all the parameters measured using R.T.

The C.S. technique, of course, requires a longer training period but the more standardized, albeit cumbersome, procedure needed to achieve pre-determined section planes and better resolution probably tend to reduce subjective operator error. This is further emphasized by the results of the measurements on the second coded copy: while there were no significant differences with HC or AC among operators using C.S. as between measurements obtained from line and from retained photographs, significant differences occurred among operators using R.T. (HC:  $p < 0.001$ , AC:  $p < 0.05$ ). Indeed, ease of measurement depends on the quality of the picture obtained, which in turn depends on the skill of the operator. An unexperienced real-time operator is prone to use only a narrow range of settings of the more simplified real time/time gain compensation, exercises less care in keeping the ultrasonic beam perpendicular to anatomical landmarks (such as "midline", umbilical vein and so on) and in avoiding compression of the mother's abdomen (which may reduce amniotic fluid around the fetal section)

and does not generally take proper advantage of acoustic windows.

Notwithstanding the greater influence of operator experience on the real time results, the mean values of BPD and AC obtained with C.S. and R.T. were not statistically different. Interaction was significant between the two machines but clinically unimportant (figure 3) and did not always occur in the same direction.

Other authors have reported that R.T. and C.S. give homogeneous results for both BPD and AC [8, 9, 19] whereas no data have been reported on HC. The need to extrapolate the perimeter of a fetal section, due to the limited width of the linear-array transducer, has already been claimed to be a possible source of error [19]: the shape of the fetal skull, which often requires this procedure, may explain the statistical difference in HC between the two machines reported in the present study.

When a new diagnostic tool has become part of common clinical practice, its reliability is still assumed to be that reported by leading research groups. Clinical results often do not coincide with expectations on the basis of reference standards. This is especially true of the ultrasound technique the use of which has spread in recent years to almost all medical centers, and involves a large number of doctors and technicians with a wide variety of clinical and specific ultrasonic experience using very different equipments. The reliability of the laboratory as a whole as assessed in this investigation can give a better estimation of the clinical results that can be achieved.

The present study, which involved six operators with different experience using two different types of machine, one of which is a real time basic scanner, allows us to draw the following conclusions:

1. the accuracy of the parameters studied does not change throughout pregnancy;
2. operator experience is more important for real time scanners than for compound scanners; this stresses the importance of proper training programs for real time operators;

3. measurements obtained by means of real time and compound scanners can be assumed to be homogeneous;
4. the reproducibility obtained in the present study compares well with data reported by pilot studies.

### Summary

Twenty patients between 14 and 28 gestational weeks were scanned on three occasions during a single day by six operators with different ultrasonic experience (three used a compound scanner and three a real-time equipment). On each occasion every operator had to measure in a "blind" manner the fetal biparietal diameter (BPD), head circumference and abdominal circumference (HC and AC); a coded copy of head and abdominal circum-

ference was subsequently measured by one of the members of the staff. Statistical analysis was carried out on the 1800 measurements, and it was concluded that the accuracy was not dissimilar from that reported by pilot studies. The reproducibility of the parameters studied did not change throughout pregnancy and real-time and compound scanner yielded homogeneous results.

**Keywords:** Fetal growth, ultrasound.

### Zusammenfassung

#### Genauigkeit bei Routine-Sonographien zur Messung des biparietalen Durchmesser, Kopf- und Bauchumfangs

Zwanzig Patientinnen mit einem Schwangerschaftsalter zwischen der 14. und 28. Woche wurden an einem einzigen Tag je dreimal von 6 Untersuchern mit unterschiedlicher Sonographie-Erfahrung geschallt (3 Untersucher benutzten einen Compound-Scanner, die anderen 3 ein Real-Time-Gerät).

Jeder Untersucher hatte ohne Kenntnis der vorangegangenen Messungen den biparietalen Durchmesser (BPD),

den Kopfumfang (HC) und den Bauchumfang (AC) zu bestimmen; eine kodierte Kopie des Kopf- und Bauchumfangs wurde später durch einen Untersucher nachgemessen.

Die 1800 Messungen wurden statistisch analysiert. Die Genauigkeit der Meßdaten lag innerhalb einer Streubreite, die auch andere Pilotstudien lieferten. Die Reproduzierbarkeit der untersuchten Parameter war unabhängig vom Schwangerschaftsalter. Real-Time- und Compound-Scanner lieferten übereinstimmende Ergebnisse.

**Schlüsselwörter:** Fetales Wachstum, Ultraschall.

### Résumé

#### Fiabilité d'un centre moyen d'échographie pour la mesure du diamètre bipariétal, de la circonférence céphalique et abdominale

Vingt patientes entre 14 et 28 semaines de gestation ont été échographiées à 3 reprises au cours de la même journée par six opérateurs ayant une expérience échographique différente (trois utilisant un appareillage complexe et trois un appareillage en temps réel). A chaque fois, chaque opérateur devait mesurer «à l'aveugle» le diamètre bipariétal (BIP), la circonférence céphalique et

la circonférence abdominale (CC at CA); ultérieurement un des membres de l'équipe mesurait une copie codée de la circonférence céphalique et abdominale.

Une analyse statistique a été réalisée sur 1800 mesures; elle a permis de conclure que la fiabilité n'était pas différente de celle rapportée dans les études pilotes, que la reproductibilité des paramètres étudiés ne changeait pas au cours de la grossesse et que l'échographie en temps réel et les appareillages complexes fournissaient des résultats homogènes.

**Mots-clés:** Croissance fœtale, ultrasons.

**Acknowledgements:** This work was supported by grant C.N.R. n° 8402228.56.155.02636. The authors are most grateful to Dr. EUGENIO FLORIS for his statistical advice.

## References

- [1] ANDERSEN, H. F., R. R. B. JOHNSON, M. L. BARCLAY, J. D. FLORA: Gestational age assessment. II. Prediction from combined clinical observations. *Am. J. Obstet. Gynecol.* 140 (1981) 770
- [2] BERGSJO, P., T. BAKKE, L. SALAMONSEN: Observer error in ultrasonic fetal cephalometry. *Acta Obstet. Gynecol. Scand.* 54 (1975) 41
- [3] CAMPBELL, S.: An improved method of fetal cephalometry by ultrasound. *J. Obstet. Gynaecol. Br. Cwlth.* 75 (1968) 568
- [4] CAMPBELL, S.: Ultrasonic fetal cephalometry during the second trimester of pregnancy. *J. Obstet. Gynaecol. Br. Cwlth.* 77 (1970) 1057
- [5] CAMPBELL, S., G. B. NEWMAN: Growth of the fetal biparietal diameter during normal pregnancy. *J. Obstet. Gynaecol. Br. Cwlth.* 78 (1971) 513
- [6] CAMPBELL, S., A. THOMS: Ultrasound measurement of the fetal head to abdomen circumference ratio in the assessment of growth retardation. *Br. J. Obstet. Gynaecol.* 84 (1977) 165
- [7] CAMPBELL, S., D. WILKIN: Ultrasonic measurement of fetal abdomen circumference in the estimation of fetal weight. *Br. J. Obstet. Gynaecol.* 82 (1975) 689
- [8] CLEMENT, D., R. SILVERMAN, D. SCOTT, J. HOBBS: Comparison of abdominal circumference measurements by real-time and B-scan techniques. *J. Clin. Ultrasound* 9 (1981) 1
- [9] COOPERBERG, P. L., T. CHERUB, V. KITE, S. AUSTIN: Biparietal diameter: a comparison of real-time and conventional B-scan techniques. *J. Clin. Ultrasound* 4 (1976) 421
- [10] DAVIES, P., R. E. RICHARDSON: Fetal cephalometry, an improved method and an evaluation. *Br. J. Obstet. Gynaecol.* 83 (1976) 105
- [11] DAVISON, J. M., T. LIND, V. FARR, T. A. WHITTINGHAM: The limitations of ultrasonic fetal cephalometry. *J. Obstet. Gynaecol. Br. Cwlth.* 80 (1973) 769
- [12] DOCKER, M. F., R. S. SETTATREE: Comparison between linear array real time ultrasonic scanning and conventional compound scanning in the measurement of the fetal biparietal diameter. *Br. J. Obstet. Gynaecol.* 84 (1977) 924
- [13] FERRAZZI, E., U. NICOLINI, A. KUSTERMANN, G. PARDI: Efficacia della attività ambulatoriale nello screening della patologia dell'accrescimento. In: DESTRO, F. (ed.): *Ultrasuoni in Ostetricia*. Monduzzi, Bologna 1981
- [14] LITTLE, D.: Reproducibility of real time ultrasound fetal measurements. In: BENNETT, M. J., S. CAMPBELL (eds.): *Real-time ultrasound in Obstetrics*. Blackwell Scientific Publications, Oxford 1980
- [15] LUNT, R., T. CHARD: Reproducibility of measurement of fetal biparietal diameter by ultrasonic cephalometry. *J. Obstet. Gynaecol. Br. Cwlth.* 81 (1974) 682
- [16] POLL, V.: Precision of ultrasonic fetal cephalometry. *Br. J. Obstet. Gynaecol.* 83 (1976) 217
- [17] SABBAGHA, R. E., M. HUGHEY, R. DEPP: Growth adjusted sonographic age: a simplified method. *Obstet. Gynecol.* 51 (1978) 383
- [18] TAMURA, R. K., R. E. SABBAGHA: Percentile ranks of sonar fetal abdominal circumference measurements. *Am. J. Obstet. Gynecol.* 138 (1980) 475
- [19] WEINER, C. P., R. E. SABBAGHA, R. K. TAMURA, S. DAL CAMPO: Sonographic abdominal circumference. Dynamic versus static imaging. *Am. J. Obstet. Gynecol.* 139 (1981) 953

Received December 20, 1984. Accepted February 25, 1985.

Dr. Umberto Nicolini  
 Clinica Ostetrico Ginecologica  
 "L. Mangiagalli"  
 Università degli Studi di Milano  
 Via Commenda 12  
 20122 Milano, Italia

# **Spina bifida – Neural Tube Defects**

Basic research, interdisciplinary diagnostics and treatment, results and prognosis

Edited by *D. Voth* and *P. Glees* in collaboration with *J. Lorber*

1986. 17 x 24 cm. Approx. 350 pages. Cloth approx. DM 148,–  
ISBN 3 11 010768 6

This book deals with, in a comprehensive yet concise form, the formal origin of dysraphic malformations and at the same time surveys the present standard of knowledge regarding research on its causes and prevention.

The special features of the epidemiology together with the basic research carried out by the Sheffield Working Group, are discussed at length. The problem of both antenatal diagnosis and pregnancy termination with the resulting direct and indirect consequences are treated extensively.

Other chapters deal with operative treatment of dysraphic malformations and the complications and results thereof.

Finally, orthopaedic specialists give extensive information on deviations of the spine (which tend to occur frequently) and its necessary therapeutic treatment.



**de Gruyter · Berlin · New York**