Introduction

The effects of low birthweight on both the quantity and the quality of the survivors of the neonatal period have been very extensively studied, and they seem to be uniformly adverse. But this information alone is of no practical help to the obstetricians and paediatricians who have to make the relevant day-to-day clinical decisions, since low birthweight can be caused by two totally different abnormalities of the intra-uterine phase of growth — either shortening or slowing (or a combination of the two). It is only these primary abnormalities whose incidence or consequences can be modified by clinical actions, so it is their specific effects which the clinician needs to understand as fully as possible.

A great deal of information has become available relatively recently concerning their perinatal effects. The causes of death and the clinical complications of the first week after birth associated with shortening of the intra-uterine phase of growth (being born too soon) are clearly different from those associated with slowing of the nett rate of intra-uterine growth (being born too small) (Butler and Bonham 1963, Neligan *et al.* 1963, Fedrick and Butler 1970). No comparable body of information is yet available about their longer-term effects on the quality of the survivors of the first weeks of life, so the obstetrician who is trying to decide whether to shorten the duration of a pregnancy by artificial means, in the interests of the baby's survival, is unable to balance the risk of long-term handicap which could result from the baby being born too soon against the risk which could result from his staying longer in a uterus where he does not appear to be growing at an adequate rate. In the same way, the paediatrician who is faced with the care of a baby who has been born too soon cannot say how important it is for his subsequent growth and development that he should achieve a good rate of growth during what should have been the latter part of the intra-uterine phase. The main purpose of the study, the results of which form the basis of this book, has been to make a contribution towards solving these dilemmas. For reasons which are discussed below, such relevant information as has yet been made available is scanty and superficial and suffers from technical inadequacies, either in the selection of the populations studied or in the methods of assessment used, or both.

There are a number of special features which characterise our study and which we believe make its results suitable for our purpose.

(1) *The population studied* was selected specifically for the purpose of comparing a group of children who had been born too soon and a group of children who had been born too small, both with each other and with a suitable comparison group, in terms of many aspects of their performance during their early school years. Our two abnormal groups represented the two abnormalities of intra-uterine growth which can account for low birthweight, and which face clinicians with different practical problems.

(2) Our two abnormal groups and our comparison group were selected from a much larger, *geographically defined population* for whom simple but standardised obstetric, paediatric and social data were available through the Newcastle Survey of Child

Development (Neligan *et al.* 1974). The selection was therefore free of the biases due to medical, social or personal factors which inevitably distort the composition of groups selected from the births in a particular hospital, and which make it difficult or impossible to apply the findings to other populations.

(3) *The methods of assessment* used were sufficiently sophisticated, comprehensive and specific to detect very minor differences between individual children over a very wide range of measures of performance. Great care was taken to enable the results to be quantified in terms of continuous variables.

(4) *The methods of statistical analysis* employed have enabled us both to identify the specific and independent effects of the two abnormalities of intra-uterine growth in which we are primarily interested (after allowing for a large number of associated, potentially confusing, factors) and to compare the magnitude of the effects of these factors.

It is instructive at this stage to consider why so many investigators studied the *sequelae* of undifferentiated low-birthweight (their published reports were well reviewed by Benton 1940 and Wiener 1962), and why it took so long for interest to move on to the much more constructive and practical study of the underlying causes. The simple answers would appear to be (*a*) that birthweight, being relatively easy to measure, could be reliably made available for every baby born; and (*b*) that before the underlying causes could be investigated it was necessary to apply to the study of the intra-uterine phase of growth the same methods of thought, and as nearly as possible the same methods of investigation, as had been applied to the postnatal phase. There are obvious inherent technical difficulties involved in trying to do so because of the very protected and inaccessible situation of the baby in the uterus, but these difficulties also were aggravated for many years by a confusing official nomenclature. The World Health Organisation had recommended in 1948 that the very vulnerable group of babies with a birthweight of 2.5kg or less should be called 'premature'. This recommendation had given official support to a confusion of ideas concerning the relationships between birthweight and gestational age which had already been causing trouble in certain circles even without such backing. An obstetrician had written: "In such circumstances, when the patient has been delivered of her baby within one or two weeks, or even after the estimated date of parturition, it is extremely annoying to have a paediatrician insist that the baby is premature because it weighs only 4½ or 5lbs" (McBurney 1947). With official support, this confusion spread and gained strength, so that it became a common experience to find an obstetrician who had delivered a mother of an unexpectedly small baby persuading her (at least to his own satisfaction) that she was wrong about her 'dates', so shortening the recorded period of gestation to correspond with his idea of what was appropriate to the baby's size.

By 1961 it was recognised how much trouble the earlier recommendation was causing, and the definition was altered to suggest that in future the same group of very vulnerable babies should be called "of low birthweight" (WHO 1961). This recommendation established the important and valuable principle of describing babies in terms of a characteristic which could be measured, rather than one which was (often misleadingly) inferred. It also liberated a flood of constructive thought by

clinicians about variations in intra-uterine growth and the factors which could determine birthweight. One of the first practical results was the publication of a set of centile charts relating birthweight to gestational age (Lubchenco *et al.* 1963).

The construction of such charts is based upon the same principles as underlie the construction of postnatal growth curves based upon cross-sectional data. In either case, sufficient observations are made during each interval of age, either gestational (calculated from the date of the mother's last menstrual period) or postnatal (calculated from the date of birth). The desired centile values for each age-interval can then be calculated, and the corresponding points for successive intervals are joined up to make the centile curves. Once appropriate norms of this type became available for the intra-uterine phase of growth, it was possible to apply to it the same methods of thought as had long been applied to the postnatal phase. For example, it no longer made any sense to treat 'weight at birth' as a variable without standardising for the calculated gestational age (of anything from about 28 to more than 42 weeks), any more than it would make sense to treat 'weight at school entry' as a variable without standardising for the calculated postnatal age (anything from perhaps three to five years). The fact that it is possible to learn something useful about the effects of variations in intra-uterine growth by using the unstandardised birthweight as a continuous variable, as has so often been done, has two different explanations. The first is that the lower centile curves relating birthweight to gestational age become more or less horizontal over the period when the majority of births occur. This flattening of the relatively crude 10th centile used in our present study can be seen to occur at about 38 weeks (see Fig. 1, p. 5). This means that after this stage of pregnancy a low birthweight implies a slow nett rate of intra-uterine growth, without the need to standardise the weight by reference to the centile curves. From data published elsewhere and based on nearly 30,000 births to Newcastle mothers (see Fig. 7.1, Neligan *et al.* 1974), it can be calculated that nearly 93 per cent of all the births occurred at or later than 38 weeks (266 days). Only about 7 per cent occurred at an earlier stage, when a similar low birthweight would need to be standardised by relating it to the baby's gestational age and the sloping position of the centile curves before its implications in terms of intra-uterine growth rate could be identified. The second explanation is that a high birthweight, above about 4.0kg, implies a rapid nett rate of intra-uterine growth, whatever the gestational age, so deductions about its implications can be made without any standardisation (Neligan *et al.* 1974, Ch. 7).

However, the nature of the data available concerning the intra-uterine phase of growth renders it inherently less satisfactory for the purpose of constructing such norms than are the data which can easily be accumulated for the postnatal phase. There are two main reasons why this is so.

(1) The true gestational age may at worst be completely unknown (because the mother's 'dates' are unknown or uncertain) and at best is less reliably and accurately known than the postnatal age. The menstrual date does not bear a constant relationship to the date of ovulation and conception (which is what we really want to know), so that variations in the interval between menstruation and conception inevitably distort or blur the picture to some extent. Moreover, even when the date of the last menstrual period is 'certain' by normal clinical standards, it appears to be

subject to considerable errors of reporting, some of which are systematic. For instance, Lubchenco *et al.* (1963), Neligan (1965) and Gruenwald (1966) all found an excess of heavy babies at all gestational ages below about 37 weeks. There are good reasons for believing that this excess is made up of full-term babies whose mothers' dates of last menstrual period have been misreported in all good faith because one or more episodes of bleeding had occurred after the start of the pregnancy. These account for about one-third of all births in the earlier gestational age-groups, and a simple method for correcting the raw data for use in constructing centile curves was suggested by Neligan (1965) and by Gruenwald (1966). The great importance of the Aberdeen curves published by Thomson *et al.* (1968) is that they are based upon the only collection of raw data derived from a large, geographically defined, population which was virtually free of this systematic distortion of pre-term birthweight distributions. This quality appears to have been achieved by the care with which the obstetricians concerned excluded from the data all mothers in whom there was a discrepancy between the date of the last menstrual period and the clinical findings at the first antenatal visit in early pregnancy.

(2) The great majority of all births occur at or after full term, so that the absolute number of observations available in the gestational age-intervals before 37 weeks is relatively very small. Altogether, only about 5 per cent of babies whose mothers have 'certain' dates are pre-term. As a result, a very large total population of births has to be covered to provide sufficient data for calculation of valid percentile values in the pre-term age-groups.

Because of these characteristics of the generally available raw data, any attempt to study the later development of pre-term babies is faced with the difficulty of deciding which of the heavier babies reported as being born before 37 weeks should be retained in the study, and which should be excluded on the grounds that they are (probably) misreported full-term babies. Our solution to this problem is described in Chapter 1 (p. 5). The better solution, of assessing each baby's gestational age clinically after birth and rejecting those where there has clearly been an error, was not available to us when we began this study since the methods had not then been described (Dubowitz *et al.* 1970, Parkin 1976). In any case they would have been difficult to apply to our large population of births, 40 per cent of whom were delivered in their own homes.

Nevertheless, in spite of its inherent difficulties, our approach to the study of variations in the duration and the rate of intra-uterine growth constitutes a clear step forward in biological and clinical terms: it was strongly supported by a working party of the second European Congress of Perinatal Medicine in 1970. Among its other benefits, the study has led to recognition of the extreme importance of establishing the true duration of every pregnancy as reliably as possible so that the obstetrician's decisions about intervening to prolong or to shorten the duration of the baby's stay in the uterus may be soundly based (Lind 1970). This approach has also made it possible for the obstetrician to begin to think about the possible long-term, as well as short-term, effects of variations in the rate of intra-uterine growth (Rhodes 1973). It has also made it possible for the paediatrician to think in a similar way about the effects of the neonatal management of babies whose low birthweight is attributable to

different causes (Davies and Davis 1970). Highly relevant experimental work has shown that undernutrition at a critical stage of development (the relationship of which to the usual time of birth differs in various animal species) may produce a permanent deficit or distortion of physical growth (Widdowson and McCance 1963), including that of the brain (Dobbing 1970). Human beings are now being studied in adequate numbers to support working hypotheses about the most vulnerable stages of development, and about the neurological functions (those of the cerebellum) most likely to be affected by adverse factors occurring during the last trimester of pregnancy and the first few months of postnatal life (Dobbing and Sands 1973). The concept of intra-uterine malnutrition, so eloquently described in clinical terms by McBurney (1947), has been established on a firm biochemical basis by Lindblad (1970). The structural effects of postnatal malnutrition upon the brains of infants have been documented by Winick (1969), and two groups of workers have taken advantage of the 'pure' malnutrition associated with hypertrophic pyloric stenosis to demonstrate that it can produce later deficits in learning ability (Klein *et al.* 1975) and in adult stature and fertility (Berglund and Rabo 1973).

Following the emergence, from about 1961 onwards, of the new approaches to the problems of low birthweight and of variations in the duration and rate of intra-uterine growth, there was inevitably a considerable delay before the new concepts could be applied satisfactorily, and on a large scale, to studying the sequelae in later childhood. Ideally, such studies require a number of conditions to be fulfilled.

The first is that appropriate norms relating birthweight to gestational age should be available for use in selecting a study population with the desired degree of variation in duration and nett rate of intra-uterine growth. Because gender is a fundamental biological characteristic which has well-marked effects upon growth and other aspects of performance both before and after birth, the norms should be sex-specific. Secondly, the study population should be selected from a population of births whose gestational ages have been recorded with as much interest and concern as their birthweights. Because of the unreliability and the tendency to error in reporting of dates of last menstrual periods, these should have been checked by a careful clinical assessment during the reported first trimester of the pregnancy to exclude those cases with an obvious discrepancy. No retrospective 'correction' of the dates should have been allowed at any later stage (to avoid the 'dates' being altered to correspond with the size of the uterus, so obscuring variations in its growth rate). The third condition is that perinatal complications which might modify the structure of the child's nervous system should have been identified and recorded. We have in mind adverse factors which could cause damage through trauma or biochemical disorders, and which might in some instances be systemically associated with abnormalities of intra-uterine growth. Fourthly, the selected children should have been followed prospectively to an age when reliable and sensitive measurements of all aspects of development can be carried out; and lastly, environmental factors which might modify the effects of the perinatal factors upon the later measurements of development should have been identified and recorded. Some of these might modify the physical structure of the children's brains (*e.g.* encephalitis or head injury), but these are likely to be rare. Much more frequently, environmental factors might modify the

children's later performance by psychological mechanisms acting from the time of birth onwards. These mechanisms might include interactions between the temperament of the child and that of his parents (Chess *et al.* 1960, Thomas *et al.* 1968); and the parental attitudes might be adversely affected by the restricted contact with the child during the neonatal period as a direct consequence of abnormalities of intra-uterine growth. Such effects may be produced by the restrictive practices of a special-care nursery (Oppé 1960), which are liable to be exaggerated if the child is nursed in an incubator (Scarr-Salapatek and Williams 1973).

It can be said at once that, to our knowledge, no study yet reported fulfils all these requirements. Our own certainly does not do so, but we do think it provides the best data yet available. But before we go on to describe the characteristics of our study we would like to discuss other relevant studies which have been reported to date. We exclude from this discussion studies such as those of Warkany *et al.* (1961) which included children with gross congenital malformations, in whom it seems likely that intra-uterine growth retardation and subsequent handicaps had a common cause. The studies with which we are concerned are those in which gross congenital abnormalities have been excluded, the assumption being that variations in the duration or the rate of intra-uterine growth (or their perinatal complications) may have a direct causal relationship to any variations in the quality of the survivors of the neonatal period. Even in such studies, of course, it is not possible to rule out the presence of a common causal factor in every case.

The first such studies of the effects of impaired intra-uterine growth took advantage of the fact that the gestational ages of the two members of a twin pair are the same, either literally or within a matter of hours, so that when there is a considerable discrepancy between their birthweights there must also be a corresponding difference between their nett rates of intra-uterine growth. Babson *et al.* (1964) and Churchill (1965) carried out studies of such twin pairs and demonstrated a definite tendency for the baby who was the lighter of the pair at birth to be relatively inferior in respect of all the physical and intellectual variables measured subsequently up to the age of 10 to 15 years. The number of children covered by these studies was relatively small, but the reports had the advantage that measurable environmental factors were the same for the two members of each twin pair. However, they suffer from the corresponding disadvantage that the performance of members of twin pairs who have grown up together are significantly different from those of singleton births, presumably because of their different emotional environments (McKeown and Record 1971).

The results of studies of singleton births are clearly of more general application but, as mentioned earlier, there are difficulties in fulfilling the ideal conditions. The special problem of bringing practical results into line with the new methods of thought, which began to be more widely disseminated from about 1961 onwards, is beautifully illustrated in the book by Drillien (1964). Although the Introduction to that book makes it clear that the author was now well aware of the theoretical and practical advantages of considering birthweight in relation to gestational age, rather than in isolation, her results were derived from a population (born in 1953 to 55) which had been selected on the basis of birthweight alone. In the rest of the book

there is only the briefest mention of the effect of variations in the relationship between birthweight and gestational age, and that only in connection with later physical growth. In a more recent report (Drillien 1970), covering the intellectual performance of the same children at the age of 10 to 12 years, the population has been subdivided in terms of birthweight for gestational-age centiles, within groups primarily defined in terms of birthweight alone. This procedure obscures any effect which may be produced by variations in the duration of gestation and makes it unnecessarily difficult to identify the effects of variation in the nett rate of intra-uterine growth. In other words, the results of such analyses fail to provide the answers to the practical, clinical questions which face the obstetrician and the paediatrician.

This same comment applies to the results of a number of other published studies in which some form of subdivision by birthweight for gestational age has been superimposed upon an initial selection in terms of 'horizontal' birthweight criteria alone (however interesting and important some of these studies may be in other respects). Examples are the studies of nearly 1000 children with a birthweight of less than 1800g (McDonald 1967); of 420 weighing less than 2040g (Eaves *et al.* 1970); of 500 weighing less than 2500g (Wiener 1970); of 202 weighing less than 2500g (Cruise 1973); of 91 weighing less than 1500g (Lubchenco *et al.* 1972); and of 105 similarly defined by Francis-Williams and Davies (1974).

The study of 96 "full term small-for-dates" infants by Fitzhardinge and Steven (1972*a, b*) fails to provide the answers for a different reason, namely that it excludes babies who were born too soon. The study by Barker (1966) has the advantage of being based upon a large, geographically defined community, and does examine the separate effects of variations in birthweight alone, of birthweight/gestational age (in a rather crude way) and of gestational age alone — but only in terms of the incidence of "non-specific mental subnormality". Barker and Edwards (1967) reported the effects of variations in the duration of gestation (but not of intra-uterine growth) upon the verbal reasoning scores of children aged 11 years, and found that both shortening and prolongation of gestation produced impairment of performance. How this compares with the impairment of performance universally reported in babies of low birthweight cannot be decided on the evidence referred to so far.

Two published reports of the findings of the Newcastle Survey of Child Development, which covered the geographically defined population from which our own study population was selected, are relevant. The first (Neligan 1967) gave some tentative answers to the practical questions, for it compared the subsequent physical growth and intellectual performance at age five years of groups subdivided in terms of gestational age alone, and in terms of birthweight for gestational age (reflecting nett intra-uterine growth rate). But intellectual performance was measured only by the Goodenough Draw-a-man Test, and the number of children available for analysis was small, so the suggested answers were very tentative. Neligan *et al.* (1974) reported results derived from all the (aproximately) 9000 children available for assessment at the ages of five and 10 years, but did not specifically address themselves to these same practical questions which form the main focus of interest of our own study. They compared the effects of variations in gestational age (which were slight) with those of variations in birthweight over the whole range (which were very significant), the

authors being more interested in identifying the beneficial effects of high birthweight (which necessarily implies a high nett rate of intra-uterine growth) than in confirming the harmful effects of low birthweight (without subdivision in terms of gestational age, so without specific reference to poor intra-uterine growth rate).

The national study by Davie *et al.* (1972) did examine the effects of variations in both the duration and the rate of intra-uterine growth, but only in terms of "educational backwardness", "poor copying designs score", "clumsiness", "social adjustment" and "recognised handicap". Both shortening and prolongation of the duration, and also slowing of the rate, appeared to have harmful effects, but the magnitude and quality of these effects were not compared.

The effects of perinatal complications have been studied in a number of groups of very low-birthweight babies, but usually only in terms of the incidence of gross handicaps (McDonald 1963, Drillien 1967, Lubchenco *et al.* 1972, Davies and Tizard 1975). Such babies have been studied in a similar way, and also in terms of cognitive function and behaviour, over the full range of birthweights in our Newcastle population (Neligan *et al.* 1974). Adverse effects have been found to be surprisingly infrequent, though their importance may be out of proportion to their incidence if they can easily be modified by changes in clinical practice. Moreover, some are systemically related to abnormalities of intra-uterine growth: examples are the relationships between preterm birth and neonatal hyperbilirubinaemia (Zuelzer and Brown 1961) and between 'intra-uterine malnutrition' and neonatal hypoglycaemia (Neligan *et al.* 1963).

Drillien (1970), in a paper specifically concerned with the 'small-for-date' infant, reported an increasing incidence of congenital anomalies with decreasing birthweight, and an association between these and subsequent handicap. But since the nature of the anomalies was not specified, and "there is also no obvious explanation why congenital anomalies should be associated with mental and neurological defects", it is difficult to assess the importance of this observation. Stewart (1972) failed to confirm it.

Many studies have emphasised the importance of attempting to allow for modifying effects of social and other biological factors which may be associated with the factors of primary interest. The associated factors which Davie *et al.* (1972) allowed for were social class (derived from the father's occupation) and the child's birth order (derived from the number of previous viable pregnancies). Wortis and Freedman (1965) studied the effects of various social factors upon the development of 'premature' children up to the age of 2½ years. Drillien (1970) found that adverse effects of being 'small-for-dates' were demonstrable at 10 to 12 years in children from "average or poor working-class" homes, but not from "superior" homes. Eaves *et al.* (1970) found that socio-economic factors began to outweigh the effects of low birthweight at about the age of four years. The effects of such factors are not, of course, confined to babies of low birthweight: for instance Illsley (1966, 1967) demonstrated the effects of maternal age and parity, as well as social class, upon the test scores of all Aberdeen children at the age of seven years. Neligan *et al.* (1974) found significant effects of the child's sex, birth rank and social class, and of the mother's standard of child care, upon the performance at the ages of five and 10 years

of the Newcastle children from whom our study population was selected.

Losses from the original population is another problem to be faced when planning a longitudinal prospective study of the later effects of factors which can be identified at birth. Fitzhardinge and Stevens (1972*a*, *b*) had lost 27 per cent of 131 small-for-dates infants by the age of four years and were unable to give any information about the development of these missing children. Wiener *et al.* (1968) were unable to test only 17 per cent of their 992 low-birthweight and control children at eight to 10 years, and they were able to state that 1.9 per cent were known to be too severely handicapped for testing and that 5.9 per cent had been examined by others at the age of five years. Therefore in only 9.2 per cent of the original population was there complete uncertainty about developmental status, but even this is worrying since there is a clear tendency for children 'lost' in this way to include an excess of the moderately or severely handicapped (Neligan *et al.* 1974, Ch. 10). The lowest loss rate we know of was reported by Drillien (1967), who retained all but 5 per cent of 104 children of very low birthweight to beyond the age of five years.

Finally, in a study of this kind there is the question of the methods of assessment to be used at all stages. This is of the greatest importance, and the dilemmas facing the investigator are discussed by Neligan and colleagues (1974) in their Introduction. In principle, there is an inverse relationship between the scope and complexity of the methods to be used in assessing each child, and the number of children whom it is practical to include in the investigation. Barker and Edwards (1967) covered a city's whole population of 50,046 children, but relied on routine clinical records of both hospital and domiciliary confinements for all antenatal and perinatal data, and assessed the children's quality at the age of 11 years by the results of two group tests of verbal reasoning. Rubin *et al.* (1973) followed 78 low-birthweight and 163 other children to the age of seven years, using batteries of tests of mental and language development and school readiness, and assessing academic achievement and school-identified problems. In the series of studies reported by Drillien (*op.cit.*) there was a high level of personal involvement, with assessments based upon a combination of subjective evaluation and the results of objective tests. The Baltimore studies reported by Wiener *et al.* (*op.cit.*) are unique in that they represent an attempt to apply extensive and complex objective methods of assessment at intervals up to the age of 10 years to a population which initially consisted of 500 low-birthweight and 492 "full term" babies; and to use sophisticated techniques of multivariate analysis in identifying the various effects in which the authors were interested.

All the problems discussed above have had to be faced in the planning, execution and analysis phases of the investigation which we describe in the rest of this book. We feel that we have met varying degrees of success in achieving our various objectives, and we have attempted to assess the quality of our conclusions accordingly.