The International Fetal and Newborn Growth Standards for the 21st Century

INTERGROWTH-21st

Study Protocol

International Fetal and Newborn Growth Consortium

The Aga Khan University Hospital, Kenya; **Aga Khan University Medical Centre**, Pakistan; **America Arias Hospital**, Cuba; **Capital Medical University**, China; **Ketkar Nursing Home**, India; **Universidade Federal de Pelotas**, Brazil; **St George's, University of London**, UK; **National Institute of Public Health**, Mexico; **Ministry of Health**, Sultanate of Oman; **University of Oxford**, UK; **Complejo Hospitalario de Toledo**, Spain; **University College Cork,** Ireland; **Department of Nutrition, World Health Organization**, Switzerland

UNIVERSITY OF OXFORD

Nuffield Department of Obstetrics and Gynaecology, University of Oxford, UK (Dr. José Villar; Dr. Stephen Kennedy);

St George's, University of London, UK (Dr. Aris Papageorghiou);

Universidade Federal de Pelotas, Pelotas, Brazil (Professor Fernando Barros);

Ketkar Nursing Home, Nagpur, India (Dr. Manorama Purwar);

Department of Paediatrics and Child Health, Aga Khan University Medical Centre, Karachi, Pakistan (Professor Zulfiqar Bhutta);

Beijing Obstetrics and Gynaecology Hospital, Maternal and Child Health Center, Capital Medical University, Beijing, China (Professor Pang Ruyan);

Ministry of Health, Muscat, Sultanate of Oman (Dr. Yasmin Jaffer);

Complejo Hospitalario de Toledo, Toledo, Spain (Dr. Anibal Nieto);

America Arias Hospital, Havana, Cuba (Professor Alejandro Velasco; Professor Ariana Valdes; Dr. Ana Chamizo);

The Aga Khan University Hospital, Nairobi; Kenya (Dr. Maria Carvalho);

Department of Epidemiology and Public Health, University College Cork, Ireland (Dr. Jan Van den Broeck);

Centre for Research in Nutrition and Public Health, National Institute of Public Health, Mexico (Dr. Juan Rivera);

Department of Nutrition, World Health Organization, Geneva, Switzerland (Dr. Mercedes de Onis)

Acknowledgements

Table of Contents

I. Brief Summary

Charitable Purpose:

To develop new "prescriptive" standards describing normal fetal growth and newborn nutritional status in eight geographically diverse populations, and to relate these standards to neonatal health risk. The worldwide use of these tools should improve infants' health and nutritional status.

Project Description:

The project aims to develop scientifically robust clinical tools to assess fetal growth and the nutritional status of newborn infants, as adjuncts to the recently produced WHO charts for children aged 0 to 5. These will be incorporated into national and international maternal and neonatal programs, and be used to monitor and evaluate maternal wellbeing, infant health and nutrition at a population level.

To achieve these objectives, primary data will be collected on a population-based sample of healthy pregnant women. The tools will describe how fetuses and newborns *should* grow in *all* countries rather than the more limited objective of past growth references which describe how they have grown at specific times and locations. They will allow for evidence-based evaluation of nutritional status at birth and measurement of the impact of preventive and treatment interventions in the community.

II. Governance of Study

Steering Committee:
- Zulfiqar Bhutta (Husein Laljee Dewraj Professor and Chairman, Department of Paediatrics and Child Health, Aga Khan University Medical Centre, Karachi, Pakistan)
- Pisake Lumbiganon (Professor of Obstetrics & Gynaecology, Convenor Thai Cochrane Network, Faculty of Medicine, Kohn Kaen University, Thailand)
- Juan A Rivera (Director of the Centre for Research in Nutrition and Health, The National Institute of Public Health, Professor of Nutrition in the School of Public Health in Mexico)
- Aris Papageorghiou (Consultant/Honorary Senior Lecturer in Obstetrics and Fetal Medicine, St George's, University of London)
- Jan Van den Broeck (Senior Lecturer in Epidemiology, University College Cork, Ireland)
- Alison Noble (Professor of Engineering Science, Institute of Biomedical Engineering, University of Oxford)
- Doug Altman (Director, Centre for Statistics in Medicine, University of Oxford)
- Mercedes de Onis (Coordinator, Growth Assessment and Surveillance, Department of Nutrition, World Health Organization, Geneva)
- Stephen Kennedy (Head of Department & Clinical Reader/Honorary Consultant, Nuffield Department of Obstetrics & Gynaecology, University of Oxford)
- Anibal Nieto (Complejo Hospitalario de Toledo, Toledo, Spain)
- Cesar G. Victoria (Professor of Epidemiology, Federal University of Pelotas, Brazil)
- José Villar (Senior Fellow in Perinatal Medicine, Nuffield Department of Obstetrics & Gynaecology, University of Oxford) Principal investigator
- Fiona Burton (Counsellor and Lay Member, Oxford)

Executive Committee:
- Doug Altman
- Zulfiqar Bhutta
- Alison Noble
- Mercedes de Onis
- Aris Papageorghiou
- Stephen Kennedy
- José Villar

Principal Investigators
- Fernando Barros (Brazil)
- Pang Ruyan (China)
- Alejandro Velasco (Cuba)
- Manorama Purwar (India)
- Maria Cavarlho (Kenya)
- Ali Jaffar Mohammed/Dr Yasmin Jaffar (Oman)
- Anibal Nieto (Spain)
- Stephen Kennedy (UK)

Scientific Advisory Committee
- Sir David Weatherall (Former Regius Professor of Medicine, University of Oxford)
- M K Bhan (Secretary, Dept of Biotechnology, Government of India, New Delhi, India)
- Cutberto Gaza (Provost and Dean of Faculties, Boston College, USA)
- Shahida Zaidi (Professor of Obstetrics & Gynaecology, Vice President of International Federation of Gynaecology & Obstetrics, Geneva)
- Ana Langer (President & CEO, Engender Health, New York, USA)
- Michael Katz (March of Dimes Senior Vice President for Research and Global Programs, Carpentier Professor Emeritus of Pediatrics and Professor Emeritus of Public Health, Columbia University, New York, USA).

Coordinating Unit
- José Villar (Principal Investigator)
- Stephen Kennedy (Study Director)
- Layla Sheikh (Project Leader)
- Hannah Knight (Research Associate)
- Ann Lambert (Project Administrator)
- Lucy Hoch (Data Manager)
- Melissa Shorten (Project Secretary)

Ultrasound Group
- Aris Papageorghiou
- Shaida Zaidi
- Laurant Salomon
- Site Ultrasonographers TBN

Anthropometric Group
- Jan Van den Broeck
- Site Anthropometrists TBN

Neonatal Group
- Zulfiqar Bhutta
- Site Neonatologists TBN

Data Management and Analysis Group
- Doug Altman
- Laima Juodvirsiene
- Site Data Entry Persons TBN

III. Credits

This protocol was prepared by José Villar, Stephen Kennedy, Paul Chamberlain, Douglas Altman and Alison Noble (University of Oxford) and Mercedes de Onis (WHO Nutrition). These authors produced a first version, between October 2006 and May 2007, based conceptually on the published WHO Multicentre Growth Reference Study (MGRS) protocol. The following people provided comments that were considered for inclusion in the initial version: Daniel Giordano, Mario Merialdi, Guillermo Carroli, as well as several participants at a WHO meeting held in June 2007 where this first version was presented by José Villar, the Principal Investigator. We thank them for their contributions. José Villar had discussions with Daniel Giordano about data collection forms and models for data management, based on available documents and previous studies; some components have been retained in the present study and his input is much appreciated.

Further important contributions were made by Cesar Victora, Fernando Barros, Michael Kramer and Robert Platt to expanded versions of the initial document that eventually constituted the grant application to the Gates Foundation. They also made important contributions to the responses provided by the applicants to the reviewers' comments. The reviewers and staff at the Gates Foundation made considerable contributions and raised a number of important issues, all of which have been incorporated into this version. We are very grateful to all the contributors and in particular to Dr Ellen Piwoz. Finally, members of the study's Steering Committee and potential investigators contributed during the preparatory phase of the study to the final version of this protocol.

Overall, this protocol is intended to extend the concept of the MGRS protocol to cover most aspects of growth during fetal and newborn life. Therefore, this document has been developed to comply as much as possible with the WHO MGRS' rationale and procedures. We thank all the investigators and the members of the WHO MGRS Coordinating Unit for allowing us to consult their study material and to use an electronic version of their document during the early preparation of our protocol.

All the individuals mentioned above, have been invited to join the Study team at various stages and in different capacities. We still welcome contributions from those few who have not yet joined us.

IV. List of Abbreviations Used

AC	Advisory Committee
ADQU	Anthropometric Quality Control Unit
BMI	Body Mass Index
BPD	Biparietal Diameter
CRL	Crown Rump Length
CU	Coordinating Unit
DMU	Data Management Unit
EC	Executive Committee
FGLS	Fetal Growth Longitudinal Study
FL	Femur Length
IDAMIS	INTERGROWTH-21st Electronic Data Management System
IUGR	Intrauterine Growth Restriction
HC	Head Circumference
LBW	Low Birth Weight
LMP	Last Menstrual Period
MGRS	WHO Multicentre Growth Reference Study
NCSS	Newborn Cross-sectional Study
NDOG	Nuffield Department of Obstetrics and Gynaecology
NICU	Neonatal Intensive Care Unit
OMFHI	Oxford Maternal and Fetal Health Institute
PPFS	Preterm Postnatal Follow-up Study
PRC	Publications Review Committee
SC	Steering Committee
SLE	Systemic Lupus Erythematosus

V. Executive Summary

At least 60% of the 4 million neonatal deaths that occur worldwide every year are associated with low birth weight (LBW), caused by intrauterine growth restriction (IUGR), preterm delivery, and genetic/chromosomal abnormalities (1), demonstrating that under-nutrition is already a leading health problem at birth. Accurate assessment of fetal growth and gestational age for timely identification and management of growth restriction, risk assessment of undernourished newborns, and monitoring LBW trends are therefore public health priorities, especially in developing countries where 98% of all neonatal deaths occur. The long-term health implications of these conditions are now well recognized for both developed and developing countries.

Traditionally, fetal nutritional status has been assessed by measurement of uterine height in pregnancy, which can be used in first level screening but needs more precision for suspected IUGR. Birth weight, as the final point of intrauterine growth, is also used but does not adequately describe the fetal growth process. Furthermore, the newborn standards presently recommended by WHO are based on a population of births from California, USA, in the 1970s and are unlikely to be a suitable international reference in the 21st century (2). This is a particular problem for preterm babies, that have different growth patterns to fetuses still *in utero,* and for whom there are no high quality postnatal growth standards, i.e. using the prescriptive approach of the new WHO infant and child growth standards (3).

In the developed world and in urban areas in most developing countries, especially if complications are suspected, fetal growth is assessed by comparing ultrasound measurements of fetal size with reference percentiles obtained from fetuses whose growth was assumed to be normal. Unfortunately, most of the ultrasound charts of size by gestational age were obtained from very small populations of fetuses in the USA or Europe, and may not be appropriate for use internationally.

These limitations have generated concern regarding the value of ultrasound in routine clinical care particularly late in pregnancy, as well as when newborn standards are used to determine the health status of populations or monitor trends and progress in reducing under-nutrition. For example, the WHO Expert Committee on "Physical Status: the use and interpretation of anthropometry" (1995) recommended the development of international standards for both fetal and newborn growth, as well as child growth (4). The latter were recently produced by WHO and are being implemented worldwide; the former are still in the research priority agenda (5). We now propose in this project to extend the WHO work by constructing a new set of international **Fetal and Newborn Growth Standards**. It will be the first, population based, large, multi-ethnic, longitudinal, fetal growth standard based on early assessment of gestational age.

Goals and Objectives

The goal is to develop new "prescriptive" standards describing normal fetal and preterm neonatal growth over time and newborn nutritional status, and to relate these to neonatal health risk.

The primary objective is to produce a set of international **Fetal and Newborn Growth Standards** (fetal growth, birth weight for gestational age and postnatal growth of preterm infants) for practical applications in clinical use and for monitoring trends in populations. We will relate the new newborn standards to neonatal morbidity and mortality to identify levels of perinatal risk.

The secondary objectives are:
a) **Clinical:** to develop a prediction model, based on multiple 2-dimensional (2D) ultrasound measurements, for estimating gestational age during mid-late pregnancy for use in populations of pregnant women without access to early/frequent antenatal care;

10   © University of Oxford

b) **Epidemiological:** to investigate in this multi-ethnic, population based sample the determinants of LBW and its components (preterm delivery, impaired fetal growth) under current healthcare conditions, and

c) **Biological:** to acquire additional 3-dimensional (3D) images to create an anatomical and growth databank of individual fetal organs as a unique source of biological information for future research.

Project Design and Implementation

The primary objective has 3 components (Activities 1-3):

1. **Fetal Growth Longitudinal Study (FGLS)** from $<14^{+0}$ weeks gestational age to birth to monitor and measure fetal growth clinically (Symphyseal-fundal-height) and by ultrasound in a healthy population.
2. **Preterm Postnatal Follow-up Study (PPFS)** of preterm infants ($>26^{+0}$ but $<37^{+0}$ weeks) in the FGLS to describe their postnatal growth pattern.
3. **Newborn Cross-sectional Study (NCSS)** of all newborns at the study centres over 12 months, obtaining anthropometric measures and neonatal morbidity and mortality rates.

Eight geographically diverse populations will participate, covering North and South America, Europe, Africa, Western Asia and the Indian Subcontinent. Each will contribute a population-based sample of healthy pregnant women expected to provide full growth potential to their fetuses, making it possible to produce prescriptive standards, similar to the WHO Child Growth Standards. Participating sites will contribute to all the components of the project.

The study populations obtained from these geographically defined areas (e.g. city or county) should have no socio-economic constraints on growth, low morbidity and perinatal mortality and adequate nutritional status. To be included, women should be non-smokers, with a normal pregnancy history, and without health problems likely to influence fetal growth or indicate a risk for pregnancy-related pathological conditions. In **FGLS**, women will be screened before 14^{+0} weeks at the time of their first antenatal visit and followed-up with standard clinical and 2D ultrasound examinations every five weeks, i.e. up to six times during pregnancy. In **PPFS**, preterm infants ($>26^{+0}$ but $<37^{+0}$ weeks) born from this sample will be followed-up during the first 8 months of life with the same protocol and set of anthropometric measures that were used in the WHO Child Growth Study. Postnatal growth will be evaluated from both delivery and conception for comparison with the corresponding *in utero* measurements. In the **NCSS** study, all newborns at the study centres, born during a fixed (e.g. 12 month) period, will have anthropometric measurements taken immediately after birth. Only babies born to women who meet the same inclusion criteria used in FGLS will be selected to construct the newborn standards. Birth weight and gestational age will also be related to neonatal morbidity and mortality outcomes to construct risk-related newborn weight for gestational age standards. Approximately 5,000 pregnancies will be included in FGLS allowing for ~ 350 preterm infants to be included in PPFS. We expect ~500 women will have to be recruited at each study site to achieve the required number of participants. In NCSS, ~50,000 newborns will be recruited, of whom we expect close to 75% will be eligible for the standards yielding ~1,800 cases with severe newborn outcomes.

Standard quality control measures will be included, as well as a unique system of random evaluation and repetition of ultrasound measurements (from stored images) to monitor validity and reliability, and continuous real time assessment of all data collected. A limited number of experienced obstetric ultrasonographers and neonatal anthropometrists will form the research teams at each study site. Neonatal anthropometric measures will be monitored and standardized centrally. All data will be entered and managed in an on-line data management system specifically developed for the study, including a system for direct transfer of blinded data from the ultrasound equipment to the database. This on-line system allows the initiation of data analysis soon after data collection is completed. Results from all

populations will be pooled (if biologically relevant differences are not observed between sites) to construct the curves for international applications using the same analytical strategy as in the WHO Child Growth Standards. We expect that extensive secondary analyses will be conducted from these data that will make major contributions to science and clinical practice in the near future (Objectives II and III) and for many years to come. This study involves minimal risk to participants, no greater than at routine antenatal or clinical examinations. Ethical approval will be sought at the appropriate levels, informed consent obtained and confidentiality guaranteed.

Monitoring, Evaluation and Dissemination

The project will run from July 2008 to July 2013. The project's implementation will be monitored following standardized practices implemented in previous large multi-centre studies by WHO and the Perinatal Research Network, according to the set of milestones described in the proposal. It is expected that by 2014: a) The newly developed fetal growth standards will be incorporated into all obstetric ultrasound equipment produced worldwide; b) Preterm infants' growth will be evaluated using the new postnatal growth curve, and c) Newborns will be assessed using the new, international birth weight for gestational age standards. Intermediate milestones are: a) Start recruitment by April 2009; b) Complete fetal sample by March 2011; c) Complete follow-up by February 2012; d) Have data set ready for analysis by July of 2012, and e) Present final standards by the end of 2013. The dissemination of the fetal, postnatal and newborn growth standards will utilize the same channels as the roll-out of the WHO Child Growth Standards and will be coordinated by the same WHO team.

Optimizing Public Health Outcomes

The new fetal and newborn standards when incorporated into health services will be unique tools for evidence-based estimation of under-nutrition at birth, maternal and perinatal programme monitoring and the clinical care of newborns worldwide, specifically in developing countries where most of these health problem exist. When fully implemented, the growth charts we propose to develop should benefit all pregnancies throughout the world because fetal growth is currently evaluated using clearly limited tools. This should provide major clinical and economic benefits. The preterm and newborn charts should be of greatest benefit in developing countries where most morbidity and mortality amongst newborns (especially growth restricted and preterm babies) occur.

VI. Goals and Objectives

A. Goals and Attributable Benefits

The **goal** is to develop new "prescriptive" standards describing normal fetal and preterm neonatal growth and newborn nutritional status, and relate these to perinatal health risk.

The anticipated **attributable benefits** of this project are that by 2014:
a) The newly developed fetal growth standards will be incorporated into all obstetric ultrasound equipment produced worldwide.
b) Preterm infants' growth will be evaluated using the new postnatal growth curves
c) All newborns worldwide will be assessed using the new, risk related, international birth weight for gestational age standards.

B. Objectives

Primary objective:
I. To produce a set of international **Fetal and Newborn Growth Standards** (fetal growth, birth weight for gestational age and postnatal growth of preterm infants) for practical applications in clinical use and for monitoring trends in populations. The newborn standards will then be related to perinatal morbidity and mortality to identify levels of risk and target interventions.

Secondary objectives:
II. To develop a prediction model, based on multiple 2D ultrasound measurements, for estimating gestational age during mid-late pregnancy for use in populations of pregnant women without access to early/frequent antenatal care.
III. To determine in this multi-ethnic sample the determinants of LBW, as well as associated conditions, e.g. preterm delivery and impaired fetal growth and their subgroups, under current healthcare conditions.
IV. To acquire additional 3D images to create an anatomical and growth databank of individual fetal organs as a unique source of biological information for future research.

To the best of our knowledge, this project is unique because of its comprehensiveness and scale (never attempted before). It has all the characteristics recommended in the literature for constructing ideal standards for international use (6, 7): it is prescriptive, longitudinal, population-based, and gestational age will be established by ultrasound early in pregnancy. We will be able to explore associated research questions, such as possible ethnic differences in fetal/newborn growth, risk factors for LBW and its subgroups. We will also be able to predict gestational age in women with poor antenatal care and simultaneously provide a set of practical tools, which can be introduced into maternal and newborn care worldwide in a short space of time.

There is considerable demand from maternal and neonatal services to WHO and professional organizations for the standards that this study will produce (5). Hence, we have designed this project in consultation with experts in the field, as well as other research groups and institutions, in such a way that it could easily complement other yet to be developed studies, thereby reducing the risk for duplication.

VII. Project Design and Implementation

Background

In April 2006, WHO released the WHO Child Growth Standards for children aged 0 and 5 which were generated by the WHO Multicentre Growth Reference Study (MGRS) (8). Two characteristics made the MGRS unique and unprecedented as a study in its field: 1) The study included populations from several countries (Brazil, Ghana, India, Norway, Oman and the USA) and 2) A prescriptive approach was used to select the study populations, i.e. only children from populations with minimal environmental constraints on growth were included. This was achieved by recruiting children of affluent and educated parents, because high education and family income have been identified as the environmental variables most likely to be associated with optimal child growth. In addition, chronic illness, failure to adhere to MGRS feeding recommendations and maternal smoking were used as exclusion criteria.

By virtue of these characteristics, the MGRS provided the strong scientific foundations for developing standards that indicate how children should grow, as opposed to previous studies that simply described actual patterns of growth at a particular time and place. Consequently, the WHO Child Growth Standards (8) are now being used worldwide to judge children's growth because they demonstrate how healthy children grow in an environment which allows them to achieve their full growth potential.

The proposed study extends the MGRS into fetal and neonatal life. It will be based on the same prescriptive approach and international representation will be ensured by including populations from several countries. The design, implementation and conduct of the study, and dissemination of the results, as well as their incorporation into clinical practice guidelines and health care policy, will build on what has been achieved with the WHO Child Growth Standards, ensuring continuity between the development and implementation at country level of pre- and postnatal growth standards.

Conceptual Issues

There are important conceptual differences between growth references and standards. References, traditionally regarded as descriptive, are used for comparing different populations, while standards are prescriptive, implying a value judgment of optimal growth to be followed by individual pregnancies. A number of developed, as well as some developing countries have their own national standards for obstetric and neonatal use. In general, it is accepted that there is some variation in growth patterns among children of different racial or ethnic groups; however, these variations are relatively small compared to the large, worldwide variation in growth related to health, nutrition and socio-economic status. Therefore, the WHO Expert Committee on Physical Status (4) argued that "for this reason, a common reference has the advantage of uniform application allowing international comparisons without losing the usefulness for local application". Such advantages out-weigh the disadvantage of not taking into account racial and ethnic variations, if actually observed. There are also practical reasons for not developing local growth standards for each country: 1) Many populations in less developed areas experience growth deficits as a result of poor health and nutrition, so that a local reference will have less screening value for the detection of health and growth disorders; 2) Significant secular changes in growth status within a relatively short period of time may render local standards less useful for clinical screening, and 3) Proper reference development is not a task that can be done easily or repeatedly, and it is very expensive to develop local standards.

In addition, as this is a longitudinal study, we will be able to study both fetal size over time (most commonly referred to as growth charts, i.e. centiles of *size* conditional on age) and actual fetal growth, expressed by the rate of growth between time periods (velocity charts).

Rationale for International population based Standards

We believe that the current use of multiple local charts presents sufficient problems to warrant preparing simple, international standards. However, measuring fetuses and newborns from a single, ethnically homogeneous community would be inappropriate, as it would fail to reflect any worldwide variation in growth. Using multiple populations from several countries would enhance the diversity in the biological characteristics - such as parental size and maternal weight gain during pregnancy - as well as the socio-cultural factors influencing fetal growth. This approach would also be politically more acceptable than constructing the standards based on data obtained from a single country.

Furthermore, there have been suggestions that the so called "customized" birth weight standards including ultrasound-derived curves of intrauterine weight (based on estimated fetal weight) are better predictors of perinatal mortality than population-based birth weight standards similar to those we plan to produce in the newborn component of the study. We do not agree with this concept as the maternal variables included in "customized" charts, such as ethnicity, parity and maternal weight are not physiological characteristics, they are intermediate variables. There is now strong evidence, using perinatal mortality as the substantive outcome, that the benefit in prediction is not a consequence of the "customization" by maternal characteristics but rather an effect of the higher birth weight distribution obtained from ultrasound derived fetal weight charts. Therefore, these adjusted charts will identify more preterm infants as growth restricted and therefore predict mortality better because of a higher cut-off and not because they fit fetal growth patterns better (9). The proposed population-based newborn standards from a well nourished, healthy population should produce results similar to those obtained by the "customized" charts, and in addition will be related to perinatal morbidity and mortality in the same study population.

Introducing the new standards to pregnancy and newborn care:

Antenatal care consists mostly of a series of screening tests of varied complexity, implemented at different levels of care that, as a group, contribute to evaluating the overall status of each pregnancy. Objective assessments of fetal growth deviations can play a major role in prenatal care, neonatal care and outcome-based research. We are not proposing to recommend routine serial ultrasound examinations for all pregnant women in either developed or developing countries (see (10) and NICE Recommendations 2008); rather that, when they have to be used because of a clinical condition, they are interpreted correctly. Our study will produce three, new, scientifically valid tools for use at different levels of care to complement other tools already in use. The new standards will facilitate the correct interpretation of ultrasound scanning at levels of care where it is already widely used e.g. referral and most urban hospitals in developing countries. It is anticipated therefore that the new charts will have a major impact on overall care: for example, they are likely to result in fewer unnecessary interventions, such as Caesarean sections. The newborn standards will be used at all levels of delivery including rural areas; the preterm standards will be used in NICUs and similar levels up to primary health care. As an additional contribution to primary health care, we will produce new, multi-ethnic, uterine fundal height charts. We believe that such a comprehensive approach is the most effective format of care for pregnancy and newborns in developing countries.

Overall Study strategy:

Our overall strategy has been to adopt a very pragmatic approach to the study. In other words, we have tried to incorporate the study procedures into routine clinical care, so as to inconvenience the mothers as little as possible and avoid disrupting service delivery. An additional advantage of this strategy is that it increases the likelihood that the study's recommendations will be implemented and introduced into real life settings.

Nevertheless, during the preparation and implementation of the study, there will be many instances when a clear-cut answer to a clinical problem will not exist. We plan, therefore, to adopt a policy of asking the Principal Investigators and other members of the Steering Committee to make decisions in such cases. This policy will broaden our knowledge base

and create a genuine feeling of team membership, which is so important in carrying out complex, multi-centre studies. The decision-making process will be supported by an independent Advisory Group, which will be asked to resolve strategic conflicts. Ultimately, however, responsibility for finalizing the Study Protocol and the Operations Manual rests with the Executive Committee.

Selection of study sites and population framework:

There are two levels in the selection process of the study population: a) the cluster level, i.e. geographical region (e.g. city or part of a city) and within this geographical area, the selection of health institutions where women attend for antenatal and delivery care and b) the individual level, i.e. women or newborns that have the characteristics to be eligible for each of the sub- studies. We will discuss here the latter strategy, as it is common to all sub-studies and leave the former for each section on the specific sub-studies.

It is not necessary and would be unrealistic, especially in developing countries, for all institutions in a study area to fulfil all selection criteria. The basic characteristic for an institution to be considered is to serve a low risk population for impaired fetal growth. The concept is that such target populations, from which the individual participants will be drawn, should have no or be at very low risk of health, environmental or socio-economic constrains on fetal and newborn growth. Definitions will be locally adapted, for identifying socioeconomic characteristics associated with unconstrained growth in these populations. In principle, these will include markers of household income, housing tenure, education, occupation and employment status using locally selected cut-off points. These variables have been recently identified as explanatory factors in the relationship between birth weight and ethnicity (11).

Specifically, from each study site we will identify health institutions that serve populations with the above characteristics. We will conduct first a census of all hospitals where deliveries take place that are classified under local definitions as "private" or "corporation" hospitals or serving the upper socio-economic sector of the selected region. This is very important for developing countries sites. In developed countries, potential centres should also serve a general population, which will mostly have low-risk characteristics. We should concentrate on large institutions (>1000 deliveries a year), as it is important to involve only a small number of hospitals. Thus, from this census we will select those hospitals that cover at least 80% of all deliveries in the target population. However, it is recognized that some sites will not have large private hospitals and so local adaptations will have to be made.

The selected hospitals will need to agree to collaborate with the University of Oxford (Study Coordinating Unit) and provide evidence of being able to implement the study protocol; in particular, being able to coordinate referrals for ultrasound scans and having a policy of confirming gestational age by early ultrasound examination in all pregnant women. They will also need to show, using the latest data available, that the population they serve is actually of low risk as defined by: a) LBW rate <10% and mean birth weight >3100g; b) located at an altitude below 1,600m; c) perinatal mortality <20 per 1000 live births; d) mothers attending antenatal care in these institutions should plan to deliver in that or a similar hospital located in the same region; e) >75% of mothers have attained an educational level greater than the locally defined cut-off point; and f) lack of known non-microbiological contamination such as pollution, radiation or any other toxic substances (this will be evaluated in collaboration with the WHO Department of Protection of the Human Environment).

From this pool of eligible institutions that have the potential to provide the required sample size for ALL components of the study (target population), the Executive Committee and its advisors, in collaboration with the local study team, will identify the final list of units that will contribute to all the different sub-studies according to their capacity to provide the sample size needed, logistics, availability of equipment and population distribution. Individual participants will be selected from such a target population following the individual criteria listed below (see Section 1.1). Sampling for the fetal growth study will consist of all women that agree to participate and meet the selection criteria from the given date for initiation of

© University of Oxford

the study until the sample is reached (500 pregnancies). For the preterm study we will take all preterm babies, with the inclusion criteria, from this cohort and for the newborn study we will take all newborns delivered at these institutions during the 12 month study period.

A. Objective I (Activities 1 – 3)
The primary objective of the international, multi-centre study will be achieved by three Activities:
1. Fetal Growth Longitudinal Study (FGLS) from <14^{+0} weeks gestation to birth, to develop the Fetal Growth Standards;
2. Preterm Postnatal Follow-up Study (PPFS) of preterm infants (>26^{+0} but <37^{+0} weeks) in the FGLS, to develop the Preterm Postnatal Growth Standards;
3. Newborn Cross-Sectional Study (NCSS) of all neonates born over 12 months in the participating centres including anthropometric measures and indicators of perinatal morbidity/mortality to develop the Newborn Birth Weight for Gestational Age Standards.

Both longitudinal and cross-sectional data will be collected at eight geographically diverse study sites, covering North and South America, Europe, Africa, Eastern, Western Asia and the Indian subcontinent, providing subpopulations from 5 major ethnic groups worldwide. The study sites have been selected from a pool of approximately 15 institutions already identified on the basis of a proven record of submitting high-quality data to clinical studies (10, 12-14). They also a) serve populations with a socio-economic status that does not restrict fetal growth (ensuring that the curves reflect the true growth potential of fetuses, without constraints related to maternal nutrition or any other social or environmental factors); b) are located at an altitude below 1,600m, and c) have a perinatal mortality rate <20/1000, a preterm delivery rate less than 15% of the target population.

1. FGLS: to develop the Fetal Growth Standards (Activity 1)
Limitations of existing intrauterine growth charts: There are three important issues relating to the construction of fetal growth standards that, in general, were not considered when the ones presently in use were produced (4). The issues are a) **longitudinal approach**: To create a standard that reflects patterns of healthy fetal growth, serial size measurements at different stages of pregnancy are required. This provides estimates of both the amount of variability in measures among fetuses in a population, as well as the amount of variability within individuals from their own average growth trajectory (15). Some early charts were built using a single measurement from each fetus based on the assumption that cross-sectional measurements from different fetuses can be used to estimate longitudinal patterns of growth. Other authors collected multiple measurements per fetus but analyzed the data in a cross-sectional manner, resulting in an underestimation of the true variance. The present study will collect and analyze data longitudinally. b) **Small sample size**: Most intrauterine standards are based on small sample sizes, which may lead to unstable estimates of percentiles, in particular at the extremes of the distribution. The extreme centiles are those that matter for the identification of growth deviations. c) **Lack of a prescriptive approach**: In the past, references have often been built based on existing practices or from specific research units' their final aim was to have population representation in particular place and time (population representativeness). This is a descriptive approach. We believe this strategy is incorrect as the underlying population often does not possess the best health and/or nutrition. For example, the observed secular increase in birth weight could reflect the tendency towards an increase in maternal body mass index (BMI) and obesity observed in most countries and not necessarily better health status of the pregnant population. Similarly, the very high Caesarean section rates in some populations (12) can affect birth weight patterns as reflected by the increase in preterm deliveries in these populations. Therefore, the proposed new standards will be built on the basis of recommended health practices (i.e. prescriptive approach): that is, defining the target population as well-nourished, healthy (both before and during pregnancy), non-

smoking, no extreme ages, adequate education and socio-economic status. The sample to be used for the new standards should be representative of this "prescriptive" population, rather than the general population.

1.1 FGLS Individual Entry Criteria

Introduction: How to select a "healthy population" or its counterpart, the risk factors associated with fetal growth restriction, is open to considerable debate. The most important factors to consider when selecting a healthy population with no obvious risk factors for IUGR or over-growth, at the first antenatal care visit early in pregnancy, are: a) achieving a balance between strict criteria for risk and external validity of the study population and b) the logistics of screening for factors that are not part of routine care or for which consensus is lacking about their effect on fetal growth.

We have reviewed the extensive literature on this topic which has tried to separate out the risk factors for preterm delivery and IUGR (16) particularly in low and middle income countries, as well as data from our own large-scale studies which have systematically explored the associations between risk factors and perinatal outcomes. In defining the eligibility criteria for the FGLS, one could exclude women with every possible risk factor for poor pregnancy outcome. However, we believe it would be preferable, in the initial screening process, to identify factors that are proven to be related to pregnancy outcomes, specifically those that are recommended as ways of identifying women who would benefit from low-risk antenatal care.

The approach is based upon our overview of the evidence of the effectiveness of antenatal care (17), which was recently up-dated (Abalos, *WHO* 2008). Similar patterns of care are recommended in the NICE/UK Clinical Guideline (2003) and by WHO for the antenatal booking visit. We will try to ensure that the approach is implemented in the participating centres in a standardized way.

All women attending the first antenatal care visit in the selected institutions with an initial evaluation of gestational age $<14^{+0}$ weeks will be screened at study entry based upon the criteria listed below. These women define a population that is likely to need only routine antenatal care in pregnancy, i.e. it is a group of clinically healthy women who can follow basic antenatal care models. Some variables have clear thresholds (e.g. urinary culture positive or negative); for others with less clear values (e.g. maternal height), we have selected a cut-off of 153 cm from the WHO Global Survey, a very large, global, multi-ethnic Maternal & Perinatal Health Study - for a detailed description of one of the 3 regions studied, see (18). This cut-off is supported by corresponding results from a number of other international studies (19, 20)

Individual Characteristics:

Recruited women must have the following characteristics at booking:

a) aged ≥18 and <35 years;
b) BMI ≥18.5 and <30 kg/m^2;
c) height ≥ 153 cm;
d) singleton pregnancy;
e) a known LMP with regular cycles (defined as a 26-30 day cycle in the previous 3 months), without hormonal contraceptive use, pregnancy or breastfeeding in the 3 months before pregnancy;
f) natural conception
g) no relevant past medical history (refer to screening form), with no need for long-term medication (including fertility treatment and over-the-counter medicines, but excluding routine iron, folate, calcium, iodine or multivitamin supplements);
h) no evidence of socio-economic constraints likely to impede fetal growth identified using local definitions of social risk;
i) no use of tobacco or recreational drugs such as cannabis in the 3 months before or after becoming pregnant;
j) no heavy alcohol use (defined as > 4 units (40ml pure alcohol) per week) since becoming pregnant;
k) no more than one miscarriage in the 2 previous consecutive pregnancies;
l) no previous baby delivered pre-term (<37 weeks) or with a birth weight <2500g or >4500g;
m) no previous neonatal or fetal death, previous baby with any congenital malformations, and no evidence in present pregnancy of congenital disease or fetal anomaly;
n) no previous pregnancy affected by pre-eclampsia/eclampsia, HELLP syndrome or a related pregnancy-associated condition;
o) no clinically significant atypical red cell alloantibodies;
p) negative urinalysis;
q) systolic blood pressure <140 mmHg and diastolic blood pressure < 90 mmHg;
r) haemoglobin ≥11 g/dl;
s) negative syphilis test and no clinical evidence of any other sexually transmitted diseases, including clinical Trichomoniasis;
t) not in an occupation with risk of exposure to chemicals or toxic substances, or very physically demanding activity to be evaluated by local standards. Also women should not be conducting vigorous or contact sports, as well as scuba diving or similar activities

Estimation of gestational age at study entry: Clearly, establishing a very precise determination of gestational age is vitally important for constructing these growth standards (in addition to clinical management) and we are prepared to screen large numbers of women, if necessary, to obtain the ideal population.

There are three possibilities for gestational age estimation early in pregnancy: dating by a) LMP alone; b) early ultrasound alone or c) LMP and ultrasound. The implications of these different methods on research findings have recently been discussed (21). Dating by LMP and ultrasound is clearly an assessment of different parameters. Although it has been suggested that before 14^{+0} weeks of gestation ultrasound may be better by an average of 2-3 days in predicting the date of delivery, in clinical practice, both are often used in combination. Based on the 2004 Birth Cohort from the Brazilian Centre, we anticipate that

7% of women will not have a reliable LMP. In other studies in this socio-economic group, the figure may be as high as 20%.

Taking all these factors into consideration, we have decided not to base gestational age solely on LMP; rather we shall use a standard gynecological definition of LMP as the first level in calculating the gestational age. Among women with a certain LMP and regular periods who are 9^{+0} to 14^{+0} weeks pregnant, we will corroborate the gestational age with an ultrasound CRL measurement, using an internationally recognized and validated chart (21). If the difference in gestational age estimation by CRL and LMP is ≤ 7 days we will consider the LMP to be reliable and take it as the true biological date. It is acknowledged that we may exclude a number of women who might otherwise have been eligible if only ultrasound were used, but the aim is to define a population with as exact a gestational age at entry as possible.

1.2 Nutritional adequacy during pregnancy

In the WHO Child Growth Study, it was assumed from the outset that children recruited from affluent populations consumed adequate complementary foods. This proved to be the case when we analyzed their complementary diets (22). In the present study, we shall adopt the same concept but in addition will develop general nutritional guidelines for pregnant/lactating women, suitable for local use, based on the best available evidence, e.g. NICE recommended, UK Food Standards Agency advice http://www.eatwell.gov.uk/agesand Stages/pregnancy/whenyrpregnanct/ for promotion amongst the participating mothers and care providers.

Routine nutritional supplements, e.g. protein or energy, will not be given because a) it is not a component of the recommended antenatal care package (10) and b) we are only planning to recruit women with adequate nutritional status. Iron-folate supplementation, will be prescribed if necessary for anemia during/after pregnancy but given routinely only if such a policy is in place in the institution. A similar position will be taken with calcium supplementation for the prevention of pre-eclampsia and preterm delivery if such a policy is in place in the hospital.

Women will be asked to indicate if they are taking nutritional supplements and this information will be recorded in the data collection form. It is not practical in a study of this size to measure adherence in any other way or to obtain individual intakes, e.g. 24 hr recall, considering the poor reliability of such instruments for individual assessment.

1.3 Pregnancy follow-up

Women in FGLS will receive standardized antenatal care (with some local variations) based on the recommended WHO package, part of which involves screening for conditions that emerge during pregnancy.

All women recruited will be followed throughout pregnancy from the time of the first visit, irrespective of the pregnancy outcome. As a general principle, the number of exclusions will be as small as possible. They are likely to be confined to fetuses with congenital abnormalities (based on a final evaluation at birth); multiple pregnancies that were not identified at recruitment; mothers diagnosed with catastrophic or very severe medical conditions which were not present at recruitment (e.g. cancer, HIV), and those with pregnancy related conditions requiring hospital admission (e.g. eclampsia or severe pre-eclampsia). Hospital admission *per se* is not a reason for exclusion: women admitted simply for "observation" will still contribute data to the fetal growth standards unless they develop one of the conditions listed above. Most sites are malaria free ecozones; thus, intermittent preventive treatment is not routinely recommended although insecticide-treated mosquito nets may be used. In other potential sites malaria is prevalent. We will adhere to local protocols and exclude from incorporation in the fetal growth standards any woman with evidence of malaria infection during the pregnancy.

For the sake of clarity, we will produce during the preparatory phase of the study, in collaboration with the Steering Committee and experts in the field, a very detailed list of those conditions, which would result in a pregnancy being excluded from the preparation of

the fetal growth standards. In any case, if any exclusion is needed it will be done only during the data analysis period and local investigators will complete the follow-up of all enrolled women irrespective or any follow-up experience. Data from preterm deliveries, even in the absence of one of the exclusion criteria, will not contribute to the fetal growth standards.

We also expect that INTERGROWTH-21st will have fewer follow-up problems than randomized trials as it does not involve an intervention; a well-educated population will be recruited, and women and their families will like some components of the study such as ultrasound pictures or close monitoring of preterm infants. Nevertheless, to ensure that the loss to follow-up remains below our predicted <5% level, we will maintain very close contact with participants; remind them about imminent visits and ultimately conduct home visits if necessary.

1.4 FGLS 2D Ultrasound Measurements

Justification: In preparing the protocol it became clear that more than 6 ultrasound examinations after the dating scan would present unwanted logistical problems and inconvenience to mothers. It was also evident that a minimum growth change has to occur between visits, to be reliably measured by ultrasound, considering the errors in the equipment and observers. Thus, even if velocity growth by unit of time (e.g. one week) for certain parameters is high during some gestational periods, the actual change may not be reliably measured. We have considered alternative spacing between measurements such as longer intervals early in pregnancy (e.g. 8 weeks) and shorter ones in later pregnancy (e.g. 4 weeks). There is evidence, however, that measures such as BPD and femur length continue to increase fairly linearly until 34 weeks without a clear period of very high velocity growth. The logistics of coordinating multiple antenatal visits at different time intervals in a large sample mitigate against adopting variable timings. Finally, as women will be recruited more or less randomly at gestational ages between 9^{+0} to 14^{+0} weeks, we will have a spread of visits throughout pregnancy, which is also advantageous in logistical terms. In a previous report, for example the Guatemalan study, 21% of participants had one scan, 50% had two scans, 24% had 3 scans, and 5% had 4 or 5 scans at varying gestational ages (23). In the Western Australia study, participants were recruited at 16-18 weeks and scanned again at approximately 24, 28, 34 and 38 weeks (24) which we believe is less satisfactory than our schedule. We have explored further this question in a large Scandinavian data set. For 5 measures, in a fixed interval of 7 weeks between measures, they did not detect any problems in modeling growth. Therefore, after the first scan between 9^{+0} to 14^{+0} weeks, we will perform scans at ~5 weekly (±1 week) intervals, i.e. 14-18, 19-23, 24-28, 29-33, 34-38 and 39-42 weeks. Thus, a woman could have scans at 10, 15, 20, 25, 30, 35 and 40 weeks or 11, 16, 21, 26, 31, 36 and 41 weeks depending upon the gestational age at recruitment and duration of pregnancy. This has the advantage of providing greater coverage of the complete gestational period and easier scheduling of the scans.

Schedule: After the dating scan, 6 further visits (for fetal biometry) will be scheduled at ~5 weekly (± 1 week) intervals (i.e. 14-18, 19-23, 24-28, 29-33, 34-38 and 39-42 weeks). Seven measurements will be taken at each visit from 14 weeks onwards: Biparietal Diameter (BPD); Occipito-Frontal Diameter (OFD); Head Circumference (HC); Transverse abdominal diameter (TAD); Anterio-posterior abdominal diameter (APAD); Abdominal Circumference (AC) and Femur Length (FL)

At each visit, the measurements will be obtained 3 times from 3 separately generated ultrasound images in a "blinded" fashion, and submitted electronically (with the associated images) to the Coordinating Unit. The BPD, OFD, HC, TAD, APAD, AC and FL images should fill at least 30% of the monitor screen. The last recorded value of each fetal measurement will be revealed after submission for clinical purposes, as per local protocols. Standardized images will be obtained for all measurements based on internationally accepted protocols (25).

For example, the BPD will be measured from the outer-outer edges of the parietal bones in a cross-sectional view of the fetal head at the level of the ventricles. The measurement should be obtained from an image with the midline echo as close as possible to the

horizontal plane with the angle of insonation of the ultrasound beam at 90°. The image should be oval containing a centrally positioned midline echo interrupted in its anterior third by the cavum septum pellucidum. The anterior walls of the lateral ventricles should be centrally placed around the midline echo and the choroid should be visible in the posterior horn in the distal hemisphere.

The CRL to gestational age conversion will be performed using the charts developed by Robinson and Fleming (21), which remain the most widely accepted reference for early gestational age estimation. This chart is, in any case, based in a cohort of women with known gestational estimation by LMP. This is a circular argument that limits the use of ultrasound alone for the estimation of gestational age. We are aware that an ultrasound expert group in the UK is reviewing the use of these charts. We are, therefore, prepared to consider changing to a new reference if one becomes available.

We also prefer not to include participants in whom fetal size is discrepant from LMP. If we use ultrasound alone to estimate gestational age and the measurement is erroneous, that error becomes built into the growth chart. This seems another circular argument that we must try to avoid. Furthermore, the ultrasound estimation of gestational age has the limitation that all fetuses with a given CRL value will have the same gestational age estimation. This lack of biological variability is a major limitation for evaluating growth. A very interesting discussion on this point can be seen in (26)

In short, we will only consider for recruitment women with a certain LMP and regular periods; from these, we will only select those women with a CRL that is equivalent to their LMP, based on the Robinson and Fleming charts. This population is at the lowest risk of negative pregnancy outcomes (27).

Additional measurements: There is a unique opportunity to complement the study by taking additional measurements, e.g. BPD using the scanner's ellipse facility; trans-cerebellar diameter; humerus, radius/ulna, tibia/fibula and fetal foot length, and Doppler measurements of the umbilical cord artery, middle cerebral artery and maternal uterine artery. However, we are concerned that introducing extra tasks could reduce the overall quality of the seven main measurements and inconvenience the participants. Therefore, the feasibility of taking additional measurements will be evaluated in consultation with each centre before a final decision is made. In general, only some centres are expected to participate in these additional studies, which are mostly descriptive and exploratory with smaller sample sizes than in the main study.

Variability of measurements: We intend to explore several factors that might explain variability, e.g. parity, BMI, gestational age, position/presentation of the fetus, liquor volume, gender and number of measurements. We shall also introduce a comprehensive system for evaluating intra- and inter-observer variability using the 2D and 3D measures, as well as a random sample of retaken measures by an external ultrasonographer in each centre.

Equipment: All the study centres will use equipment with the same characteristics (determined by an independent group), including a state-of-the-art scanner. The staff will be appropriately trained following standardized procedures. The study will provide the equipment free to all the centres in developing countries; centres in developed countries will be expected to provide equivalent equipment themselves, as their contribution to the study. It is very important that ultrasound probes provide an extremely high quality grey scale image. We like to stress that the results of the study will be in the public domain and all manufacturers will be able to use the results free of charge. A detailed description of the methodology for ultrasound measures is presented in Appendix C.

1.5 FGLS Quality Control Measures

Quality control measures will include a unique system of random evaluation and repetition of ultrasound measurements and continuous real time assessment of all data collected. A limited number of experienced obstetric ultrasonographers and neonatal anthropometrists will form the research teams at each study site. They will be provided with standardized equipment and an operations manual that will describe all measurement techniques, protocols and procedures for training and clinical use in all the centres. The ultrasonographers will receive specific training under the supervision of qualified instructors. Intra-observer and inter-observer measurement errors will be assessed during the training course before initiating the study.

Both the examiner and the instructor will scan each fetus briefly. Each examiner will obtain two images of each fetal anatomical parameter under study. Differences between these measurements will be expressed as the percentage of the measurement obtained from the technically better image. Percentage differences will be used to take into account the increase in the dimensions of the fetal anatomical parameters with advancing gestational age. Percentage differences for each examiner will be averaged and the mean values compared to zero and to the instructor's measurements by t-test. In addition, measurement error will be evaluated by defining a maximum acceptable discrepancy, e.g. 5%, and requiring that 95% of differences lie within that value. This will provide information on the accuracy of individual measurements. The same model of ultrasound machine will be used at all sites, and the equipment will be serviced periodically by specialized technicians as per the manufacturer's recommendations.

1.6 FGLS 3-Dimensional Ultrasound Volumes

3-Dimensional (3D) ultrasound technology, a relatively new scanning technique in obstetrics, offers exciting possibilities to assess the fetus. The technique is based upon volume acquisition using specially designed probes with later, off-line review and "3D reconstruction" on a computer. This allows examination of multiple planes of the imaged structure from 3 different right angles (top-to-bottom; front-to-back and side-to-side). Additionally, a variety of techniques for measuring fetal organ and limb volumes have been described. The multiplanar display allows a clearer and more comprehensive assessment of fetal anatomy, and volume calculation produces more accurate estimates of fetal size and the impact of growth restriction on specific fetal organs (28).

The primary purpose of the 3D component in this study is quality control, i.e. to provide volumetric data to assess the quality of the 2D measures (BPD, OFD, HC, TAD, APAD, AC and FL). Hence, the 3D volumes will be the head (at the level of the BPD), the abdomen (at the level of the AC) and the femur. The task of controlling quality with these data will carried out by an independent group. The technique of 3D volume acquisition is similar for all structures. Initially, the structure to be imaged is identified during the last of the three 2D scans taken. Once imaged correctly, the volume acquisition facility is engaged and the volume obtained over a brief (10-15 seconds) interval and stored in digital form on computer for later analysis, which will be performed in Oxford by the Noble research group after the stored images have been transferred there electronically.

1.7   Criteria for excluding participants from preparation of fetal growth standards

Women in FGLS will receive standardized antenatal care (with some local variations) based on the recommended packages, part of which involves screening for conditions that emerge during pregnancy and is presently used in the hospitals. All women recruited will be followed throughout pregnancy from the time of the first visit irrespective of the pregnancy outcome, except for those with a late diagnosis of multiple pregnancy. However, we do not expect to include all pregnancies in the preparation of the fetal growth standards as some of them may develop conditions that can affect fetal growth. As a general principle, the number of exclusions will be as small as possible. They are likely to be confined to fetuses with congenital abnormalities (based on a final evaluation at birth); severe IUGR requiring early delivery; multiple pregnancies that were not identified at recruitment; mothers

diagnosed with catastrophic or very severe medical conditions which were not present at recruitment (e.g. cancer, HIV), those with pregnancy related conditions requiring hospital admission (e.g. eclampsia, placental abruption or severe pre-eclampsia) or those with acute conditions such as an episode of malaria infection. Hospital admission *per se* is not a reason for exclusion: women admitted simply for observation will still contribute data to the fetal growth standards unless they develop one of the conditions listed above.

For the sake of clarity, we will produce during the preparatory phase of the study, in collaboration with the Steering Committee and experts in the field, a very detailed list of those conditions, which would result in a pregnancy being excluded from the preparation of the fetal growth standards. Data from preterm deliveries, in the absence of one of the exclusion criteria, will contribute to the fetal growth standards until the time of delivery.

1.8   Symphyseal-fundal height charts

We will produce, in addition to the ultrasound charts, new data on symphyseal-fundal height measurements to update the charts we first produced 30 years ago based on a single sample (29). To produce one, standardized, well-validated, multi-ethnic chart using this selected population to replace several used in primary health care units and hospitals around the world would be extremely useful. We expect that these measures be taken at the same time that the ultrasound 2D measures following the same schedule by specifically trained staff.

2. PPFS: to develop the Preterm Postnatal Growth Standards (Activity 2)

2.1 PPFS Individual Entry Criteria and Follow-up

All preterm newborns ($>26^{+0}<37^{+0}$ weeks) from the FGLS cohort (anticipating ~4% neonatal mortality in this group) will be followed-up for 8 months after delivery to evaluate postnatal growth. This follow-up period means it will be possible to study all preterm newborns, even those born at 27-30 weeks, up to the ideal "term" gestational age, plus 3 full months of "true" postnatal life. We plan not to include preterm newborns ≤26 weeks (only ~3% of all preterm newborns) because of the very high morbidity/mortality in this group and need for very special intensive care. A fixed follow-up period was chosen to simplify organization of the study and reduce loss to follow-up. Nevertheless, an analysis based on the time from conception (corrected age) will be performed to compare preterm babies with their *in utero* counterparts conceived around the same time. This length of follow-up should minimize a) inconvenience to the mothers; b) the need for home visits, and c) the study's cost and complexity. An analysis based on the time from conception (corrected age) will be performed to compare preterm babies with their *in utero* counterparts conceived around the same time. Although we want to produce growth charts for the first 6 months of life, we have extended the follow-up period to 8 months for these infants to avoid the so-called "right-edge" effect in the construction of the growth standards.

Exploratory sub-group analyses will also be conducted, e.g. very preterm infants <32 weeks (~20% of all preterm newborns) and induced for medical reasons vs. spontaneous preterm deliveries. However, to be included in the PPFS growth charts, newborns have to be clinically stable and able to sustain life independently within 72 hours of birth. A detailed protocol describing clinical conditions for inclusion in the final growth curves will be prepared in consultation with expert neonatologists before any data analysis is undertaken.

This protocol will focus on definitions that can be made operational across institutions. Newborns will be excluded if they have congenital malformations, medical problems interfering with enteral feeding or need for artificial ventilation. Additional exclusions after enrolment may include any morbidities severe enough to have a negative impact on growth, feeding or the feasibility of measuring the infants at the required times. (See also section 2.4).

Clearly, for some subgroups of very preterm infants this will represent only an exploratory analysis; however, it will still be done because of the unique opportunity to study continuous fetal-preterm growth. It is expected that many babies, especially those born <32

weeks, will be managed in some form of neonatal intensive care unit. Measurements of these babies will be taken using the study protocol, but in accordance with the clinical status of the infant and the unit's protocols. It is clear that to ensure that "clinically stable" babies are comparable across different NICUs, standardization and continuous monitoring are needed. Ultimately, in such an acute clinical care setting, the final judgment will depend on the attending doctor; realistically, this is how the standards will be used in practice anyway.

2.2 PPFS Anthropometric Measurements

The anthropometric measurements will include weight, length and head circumference. The three measurements (and a standard clinical evaluation) will be taken every 2 weeks during the first 8 weeks, and then monthly until 8 postnatal months, using essentially the same methodology and operating manual employed in the WHO Child Growth Study (6). The only difference will be that in the proposed study all measures, interviews and clinical evaluations will be conducted at a special follow-up clinic in the corresponding hospital. Routine home visits are not planned, except for those mothers who do not comply with the protocol's scheduled visits. It is expected that the preterm babies will have a maximum of 11 follow-up visits over 8 months, but this number may be reduced for some as complications, including death, are expected especially in the very preterm subgroup. Abdominal circumference will not be included as it is not used in routine neonatal practice and respiratory movements in these tiny newborns make the measurements unreliable.

Standardized, electronic, digital, newborn weighing scales will be used and serviced regularly; they will be replaced if they are faulty and cannot be repaired. All anthropometrists in the study will be trained centrally; they in turn will train the nurses/midwives in how to use the scales. We will follow the same procedures used in MGRS to certify all staff responsible for taking the measurements, as was the case in our own previous randomized trials. During routine measurement sessions, two anthropometrists will independently take duplicate measurements. Maximum allowable differences between duplicate measurement values will be checked to detect outliers and trigger immediate re-measurement if necessary.

2.3 PPFS Quality Control Measures

An Anthropometric Data Quality Unit, organized on similar lines to the Ultrasound Quality Control Unit, will be created. Its role will be to conduct, following the WHO MGRS's training procedures, the standardization, monitoring, site visits, and retraining when necessary of all staff responsible for anthropometric measurements (MGRS Manual of Operation, WHO Geneva). Using the on-line data management system, samples of measurements will be obtained to evaluate variability and consistency within each site and across sites. The same protocol will be followed for measures at birth for all newborns in the fetal and newborn studies. For the preterm follow-up we will use the MGRS protocol but include only weight, length and head circumference. We will assign 2 staff at each centre to take all newborn (and postnatal follow-up) anthropometric measurements. This will permit replication of the WHO MGRS measurement standardization protocol (6).

As stated above, the standardization of all anthropometric measurements is a central element and challenge of this protocol. The measurers (or lead measurer from each site) will be brought together at a central location for two initial training meetings to be conducted by experienced staff from the WHO-MGRS team with the assistance of the MGRS anthropometry video. This video will be used for local training as well as for on-going standardization activities. The study will provide all sites with a new infantometer(s) with digital reading for infant length measures. For birth weight, hospitals will use the balances they are currently using but they will be calibrated daily. Other equipment will be calibrated at least once a week during data collection. The overall sample will be also be monitored daily, using the on-line data entry system, for unexpected values and outliers so that measurements can then be repeated within a few days, if appropriate.

2.4 Infant Morbidity During Follow-Up

For the Preterm Follow-up Study, we will use an indicator of morbidity at entry and an extended version for the follow-up itself. The protocol states (see section 2.1) that the study will exclude newborns ≤26^{+0} weeks (only ~3% of all preterm newborns) because of the very high morbidity/mortality in this group and need for very special intensive care. In addition, newborns have to be clinically stable and able to sustain life independently after 72 hours to be included in cohort from which the final growth curves will be produced. A detailed protocol describing clinical conditions, "minimum criteria", for inclusion in the final growth curves will be prepared by the study's Neonatal Group before any data analysis is undertaken. Babies that do not meet the "minimum criteria" for inclusion after 72 hours will nevertheless be followed up to form a subgroup for comparative analysis with the eligible infants. Efforts will therefore be made to standardize the major components of infant care throughout the participating institutions. The protocol will focus on definitions that can be made operational across institutions, e.g. days of IV feeding, rather than the diagnosis that led to the interruption of enteral feeding. At this stage, newborns will be excluded if they have congenital malformations, medical problems interfering with enteral feeding (e.g. a need to stop feeding for more than 5 days) or a need for artificial ventilation. Additional exclusions after enrolment will be kept to a minimum but they are likely to include any morbidity severe enough to have a negative impact on growth, feeding or the feasibility of measuring the infants at the required times.

2.5 Infant Feeding Practices

We expect newborn feeding to be protocol driven and based on a) the recent WHO recommendations (Edmond and Bahl "Optimal feeding of low-birth-weight infants" WHO, 2006) and b) standard NICU practice for clinical status. Descriptive data, collected in an adapted version of the WHO MGRS form, will serve to document that the centres conformed to recommended feeding patterns and clinical practice. The comparison with intrauterine charts is an important, biological secondary objective, but the main objective of the PPFS is to describe the postpartum growth that can be expected, or aspired, to by "healthy preterm newborns".

It is anticipated that all the sites will adhere to internationally recommended standards for feeding preterm infants; see review (30). The recommended first choice for LBW infants is breastfeeding or mother's own expressed milk. This is associated with the lowest incidence of infection and necrotizing enterocolitis, and best outcomes in terms of neurodevelopment. Where that is not possible, the next best choice is donor human milk and, in third place, preterm infant formula (Edmond and Bahl "Optimal feeding of low-birth-weight infants" WHO, 2006).

3 NCSS: to develop Newborn Birth Weight for Gestational Age Standards (Activity 3)

3.1 NCSS Individual Entry Criteria

The Newborn Birth Weight for Gestational Age Standards will be constructed from the birth weights of a) all babies born to mothers in FGLS and b) all babies born at the study centres during a fixed 12 month period (the precise period might vary depending on the total number of deliveries per year at each centre) to mothers who meet all the FGLS inclusion criteria. Thus, babies born to mothers without health problems likely to influence fetal growth will be selected from the total number of births at each centre. Their gestational age will be confirmed by an early ultrasound as in FGLS, because all the hospitals included in INTERGROWTH-21st sites, should already have a policy of checking gestational age with an early dating scan. Measures will be taken during the study period to reinforce this policy to ensure that all newborns have their gestational age at birth corroborated by ultrasound scan estimation early in pregnancy. However, the total population of newborns during the same study period will be used to explore Objective III (determinants of LBW and its components, preterm delivery and impaired fetal growth, under current healthcare conditions). The strategy of including the complete newborn population will allow us to study a large number

© University of Oxford

of LBW infants and subgroups under a wide range of conditions that can affect these outcomes. We shall also be able to compare the total populations across centres in terms of socio-demographics and pregnancy outcomes to confirm that the populations contributing to the new standards have the required characteristics.

3.2 NCSS Anthropometric Measurements

All babies born during the study period will have these anthropometric measures taken within 24 hours of delivery: weight, length and head circumference. To select those newborns eligible for the NCSS (newborn standards), we will use the same criteria as for the fetal growth study.

Standardized, electronic, digital, newborn weighing scales with a precision of 10 grams will be used and their calibration status will be checked daily; they will be replaced if they are faulty and cannot be repaired. We shall also provide all clinics with standardized infantometers for length (precision 0.1 cm) and tape measures for head circumference (precision 0.1 cm); these will be similarly calibrated and maintained. All anthropometrists will be trained centrally and monitored during the study following standard procedures by the Anthropometric Standardization Unit; they in turn will train the nurses/midwives in how to apply the study's measurement protocol. We will follow the same procedures used in MGRS to certify before and during the study all staff responsible for taking the measurements, as was the case in our own previous randomized trials. As in MGRS, routine data collection will comprise duplicate measurements of weight, length and head circumference taken independently by two anthropometrists. They will check maximum allowable differences between their replicate measurement values immediately the measurements. These maximum allowable differences are approximately twice the Technical Error of Measurement of a well-trained observer (e.g. 0.7 cm for length and 0.5 cm for head circumference). If the maximum allowable difference is exceeded, the anthropometrists will immediately go on taking other duplicate measurements until the difference is acceptable. Replicate measurement values will be averaged before analysis to increase accuracy further.

3.3 NCSS Quality Control Measures

The same quality control measures and standardization strategy (adapted from the WHO Child Growth Study protocol) (7) which are described above for PPFS will be employed for the three measurements to be taken for the whole study population in NCSS. A special Anthropometric Data Quality Control Unit, independent of the Study Coordinating Unit, will be responsible for the monitoring and quality control of these measurements during the entire study.

3.4 Follow-up of the NCSS

All newborns during the study period, including those on NICU or special care, will be followed on a daily basis until hospital discharge to detect neonatal death and document severe morbidity. We will make strenuous efforts to coordinate and promote evidence-based care for the preterm neonates using materials developed as part of the WHO best practice programme, by liaising with the lead neonatologist in each NICU before and during the study. We recognize that differences in practice will persist despite our best efforts. However, we believe this is unavoidable in a very pragmatic study such as this, which is trying to reflect what happens on a daily basis in clinical practice. Furthermore, we will similarly make strenuous efforts to standardize the main protocols for feeding practices in each NICU before the study starts. During the routine site-visits by members of the Study Coordinating Unit and the Anthropometric Team we will monitor the implementation of the protocols.

3.5 Biological Samples from NCSS

We have debated the issue of anonymized genetic/epigenetic data at great length. In principle, we would not wish to miss an opportunity to put together what would be a unique databank. However, we are very concerned about a) the effect that a request to collect DNA

might have on recruitment rates; b) ethical questions that might be raised by IRBs; c) the additional time required to obtain informed consent, and d) the extra cost of processing the samples. Having said this, it nonetheless remains an important issue. Therefore, we shall bring together a group of international experts, WHO and potential academic partners, to discuss the advantages and disadvantages of collecting biological samples for a specific hypothesis relating to fetal growth and/or preterm delivery. The final decision will depend largely on whether the study has sufficient power to answer a specific question, relating for example to the effect of a genetic polymorphism on a pregnancy outcome.

3.6 Severe Perinatal Morbidity and Mortality Outcomes

Birth weight for gestational age standards will be related to indicators of perinatal outcome to establish risk levels associated with different growth patterns. The 'ideal' outcome is perinatal mortality, but its anticipated infrequent occurrence in this low risk population makes it unrealistic to have a sample large enough for the necessary number of events across the gestational age distribution (perinatal mortality is expected to be close to 1% in these populations). We have therefore decided to use an un-weighted composite outcome including at least one of the following conditions: stillbirth, neonatal death until hospital discharge of the newborn, newborn stay in NICU for ≥7 days or other severe neonatal complications. We have used such an outcome recently (12, 31); it requires limited standardization of clinical diagnoses across hospitals and is well accepted as a marker in large, international, population based studies of newborns that are severely ill. It could be argued, however, that intrapartum stillbirth may not be related to fetal growth and should not be included in this index. We believe this is a valid point but as it will not be possible to separate those intrapartum deaths that are related to IUGR from those that are unrelated, we suggest keeping the index as it is. As we expect a very low fetal death rate, it should not have a major impact on the observed rates.

We believe this is a good proxy for adverse perinatal outcomes across countries. We have used it as a primary neonatal outcome in recent publications and it has been well accepted. Its only disadvantage is that it risks excluding from the total number of early neonatal deaths some cases amongst healthy, mostly term babies delivered vaginally who, after hospital discharge at 48 hours, develop severe complications or death up to 7 days post-natally without returning to the same hospital. However, missing these isolated cases is preferable to performing thousands of unnecessary home visits.

Data from our population-based studies in some of the sites indicate that the incidence of this outcome is close to 5%. We will confirm this estimate during the preparatory phase of the study in the study centres. We will also reach consensus with the investigators and consultants on the few very severe clinical conditions to be included in the composite outcome in addition to perinatal death and ≥7 days NICU stay.

4. On-line Data Management and Statistical Analysis

All data will be entered into an on-line data management system specifically developed for the study; it will include a system for direct transfer of blinded data from the ultrasound machines to the database. This on-line system has the practical benefit of allowing on-going quality control, correction of errors or missing values and the initiation of data analysis soon after data collection is completed. We have previously used such a system in very large multi-centre studies (12) and randomized clinical trials in developing countries, and we are confident that there will be few problems in adapting this concept to the proposed study. The system will be field tested during 2008 using the data flow model shown in Appendix B.

This system will be used for data management and monitoring all sub-studies, including patient recruitment and follow-up, and is based on the INTERGROWTH-21st Electronic Data Management System (IDAMIS). The system permits all participants' data to be incorporated contemporaneously into the data files via the Internet. Included within the system is a review process to ensure that all data are complete, and that the research team is notified about imminent patient visits and expected delivery dates. At an institutional level, each centre will have its own progress enrolment charts and weekly recruitment targets. The

system will also provide the Data Coordinating Unit with a detailed daily record of patient enrolment and data entry, at both individual and institutional levels to monitor progress against the milestones listed in the protocol. Corresponding actions, such as telephone calls, web conferences or site visits will take place within a week of detecting a problem in a centre to ensure that appropriate corrective measures are introduced. Lastly, the system will allow the database to be analyzed as soon the data collection period has ended.

Data from all sites will be pooled to construct the curves for international applications using the WHO Child Growth Study analytical strategy (8). Data analysis will be conducted following the same strategy used in the construction of the WHO MGRS curves during Years 3-5. It will be coordinated by the Statistical Director (D.Altman), consultant statistician/epidemiologist and the team's senior investigators. Extensive secondary analyses of these data will be conducted including the development of velocity standards, thereby contributing in a major way to answering questions relating to clinical practice and public health now and for years to come.

The appropriateness of pooling data from all sites will be assessed by comparisons of site means, standard deviations and the 3rd and 97th centiles to the pooled values of primary data. Consistent differences ≥ 0.5 SD between the mean values of individual sites and the pooled sample will be used as a pre-set trigger for considering whether to adjust by site for purposes of pooling data. This has to be supported by site-specific consistent differences across primary measures. For the WHO-MGRS it was concluded that data from all sites were remarkably similar and thus should be pooled (8). This is the strongest biological argument to date and will be the basis for the conceptualization of our analysis.

The planned approach will be based on the experience of the WHO-MGRS adapted to the fetal growth parameters (32); we shall follow the same exploratory strategy used in this study. We are aware that these analyses may not be automatically applicable, but we feel comfortable that the considerable experience of this team of statisticians in carrying out a similar study will enable them to select a final model that is fit for purpose. The methods do not require similar intervals. We are not making any assumptions about the shapes of the individual curves, except that they will be smooth.

For the physical growth standards, a class growth distribution that can fit the data for all measurements will be explored with an appropriate smoothing technique to generate centiles. Candidate distributions (e.g. Box-Cox t, modulus-exponential-normal, Box-Cox-power-exponential, Johnson's S_U or Johnson's S_B) will be compared on the basis of goodness of fit for each age group using the following diagnostic tools: (a) Log-likelihood comparisons, (b) Q-Q plots (i.e. comparing empirical and distributional centiles), and (c) Comparisons of observed percentages that occur below estimated centiles against expected values. Once the most appropriate distribution(s) for our data is (are) identified, smoothing techniques will be tested against the same diagnostic criteria applied to the selection of distributions such as polynomial smoothing, the Box-Cox t distribution that smoothes using natural splines and a variation of it with the Box-Cox-power-exponential distribution. Sir David Cox (University of Oxford Statistics Department) who advised us on another pregnancy-based study, will be invited to contribute to the selection process.

It could be argued that only parous women should be included in the construction of the standards. Furthermore, it would reduce the incidence of LBW babies amongst the mothers recruited as nulliparous women are known to be at higher risk of LBW babies. However, it is also the case that long-term outcomes in newborns born to parous versus nulliparous women have similar overall morbidity and mortality rates. Overall, we believe that the advantages of recruiting solely parous women are outweighed by the loss in external validity and credibility in doing so. Therefore, we intend including all women regardless of parity but we will try to recruit similar proportions across the study sites. We will also explore in the analysis if, in this sample of healthy women, parity is an important effect modifier for fetal growth, but we do not expect to produce parity-specific fetal growth standards. In addition, based on the same biological principles, sex-specific charts should be constructed for fetuses as they have been for newborns. However, we do not plan to separate fetuses by sex as this is not a routine practice and it is unlikely that it will became one in the future.

Conversely, separate sex-specific charts will be developed to monitor the postnatal growth of preterm babies.

5. Sample Sizes for Objective I (Activities 1 - 3)

General principles: The precision of growth chart centiles is determined by several factors, of which the most important is sample size. Other factors are also relevant, including study design (cross-sectional vs. longitudinal), the timing of measurements, and the method of curve-fitting. Many criteria can be used to estimate sample size, but three are considered critical here, namely the precision of: a) a given centile at a particular gestational age; b) the logistics of selecting and following-up a large cohort, and c) the rate of the expected perinatal outcome for Activity 3, in Objective I. (The outcomes to be used in Objectives II-III are either continuous variables or far more prevalent than severe neonatal morbidity/mortality, i.e. a preterm delivery rate of 9-10%, and therefore sample size limitations are of less concern for these parts of the study). Sample sizes have been estimated taking each of these criteria into account. The numbers quoted below are combined figures for all study sites.

It is possible that the WHO-MGRS findings of similar infant growth patterns amongst different ethnic groups may not apply to fetal growth, even though recent data for US white/black newborns support the concept of pathological rather than physiological factors being responsible for observed variations (16). We question why fetuses should grow differently to infants or be affected differently by environmental influences, or why biologically different growth patterns should exist across ethnic groups. Nevertheless, we will make all practical efforts to obtain the largest possible sample size per ethnic group to allow separate evaluation in case the alternative hypothesis (different fetal growth by ethnic group) is observed. Even if this is the case for some groups it is very unlikely that all ethnic groups will have biologically relevant different growth patterns.

FGLS: To obtain complete data from 4,000 pregnancies at 8 study sites, 500 mothers will have to be enrolled at each centre. Data from previous years will be used at each study site to estimate the expected number of eligible women available and the recruitment rates. We expect that ~75% of the total pregnant population at the sites selected will be eligible for this cohort, although local investigators will be strongly advised to be conservative in their estimates of compliance rates and to allow for attrition due to other reasons. The 4,000 women should provide close to 800 cases of each of the 5 general ethnic groups. Sample size calculations (15), based on a width of the 95% CI of the 5th birth weight percentile of 2.5% at term (2,700g), demonstrate that 600 women per ethnic group should be adequate, after excluding complicated pregnancies and women lost to follow-up. This sample size is larger than most previous studies and adequate, we believe, to produce reliable curves. However, we will not produce ethnic-specific standards because such pregnant populations are rarely representative of a unique ethnic group, particularly in urban areas in developing and developed countries. It would therefore be impractical to alternate between standards based on each patient's ethnic classification, even if this could be done. We will not be able to estimate outcomes across groups with any degree of power but that is not an objective of this study. These numbers should also fulfil the requirements for PPFS of ~360 preterm infants (assuming a preterm rate of 9% based on our previous study involving similar populations (12)).

We estimated that fewer than 5% of women will be lost to follow up (the figure is usually ~ 3% in our large trials). We also adjusted for the fact that 10-15% of women will be excluded from the preparation of the fetal growth standards because they will have developed problems severe enough to have affected fetal growth.

PPFS: A cohort of ~360 preterm babies will be recruited from FGLS for this study. It is expected that the gestational age distribution will provide 310 babies >30 weeks gestation. Only the subgroup of babies free of major clinical problems will be used to create the standards. We recognize that the sample size calculation here is influenced by logistic issues and the availability of newborns from FGLS rather than statistical calculations. However, it is still large by preterm study standards and we shall have very detailed follow-

up data. We consider that the possibility of having a full set of fetal and newborn growth patterns from a cohort of preterm newborns is a biological priority even if we shall not have the power to explore other subgroup analyses such as gestational age sub-groups or early postnatal morbidity. Postnatal growth from healthy populations has been shown to be similar among ethnic groups (33, 34).

NCSS: The sample size calculations for this component of the study depend mostly on the number of preterm newborns needed at the lower end of the gestational age distribution and the perinatal outcome selected. A fixed data collection period (12 months) is planned for all hospitals but it is clear that some variability will have to be accepted considering the size of each institution and the population characteristics. From the pool of all newborns, we will select the ones born to eligible women using the FGLS criteria listed above. We expect the refusal rate to be minimal. We will aim for a total of 50,000 newborns with very detailed information which will provide, based on data from similar institution (33), a sample of ~75% eligible babies for the newborn charts, including ~1,800 with a severe outcome using our established criteria. Of these 1,800, we expect that 70% (about 1,200) will be between 26 and 36 completed weeks of gestation.

Missing values: We consider that women without newborn data cannot contribute to any of the three studies. Fetal deaths and newborns with congenital malformation will be excluded. All women should have at least 2 follow-up values after the dating scan to be included in the fetal growth study.

B. Objective II: Activities 4 - 6
Objective II will be achieved by the following three activities:

1) Preparation of computer files of 2D ultrasound measures for 3-week gestational windows between 20-34 weeks;

2) Development of analytical strategies for the construction of predictive models to estimate gestational age during mid-late pregnancy and

3) Completion of prediction model to estimate gestational age during mid-late pregnancy based on several ultrasound measures obtained at a single visit. Ultrasound-based gestational age estimation in infrequent attendees or women with limited access to care is usually based on a single ultrasound measure (e.g. BPD), but the estimates have large errors and wide confidence intervals. We are planning to develop equations that use several ultrasound measures obtained at a single visit. If we achieve this, it will be a major contribution to the care of high-risk women attending hospitals on only one or two antenatal visits late in pregnancy.

Statistical strategies for the production of the prediction models include multiple regression analysis using gestational age at delivery as the gold standard (dependent variable), then recalculating the expected gestational age at different time points. The independent variables or predictors will be the basic ultrasound measures at a given gestational age window, as well as other clinical parameters that are available to the staff during the antenatal care.

C. Objective III: Determinants of LBW and its components, preterm delivery and impaired fetal growth, under current healthcare conditions (Activity 7)
The complex interactions between risk factors, clinical presentations and underlying biological processes are poorly understood in relation to adverse perinatal outcomes, especially LBW which is such a heterogeneous entity. This has been a major limitation in preventing preterm delivery and impaired fetal growth. We therefore aim to explore risk factors and perinatal outcomes for LBW (<2,500g), preterm delivery and impaired fetal growth in the entire NCSS study population of 50,000 newborns. For the identification of risk factors, data collected during routine and specialist antenatal care from ALL women delivering in the institutions over a fixed period will be included in the study. A summary antenatal and delivery form will be introduced into these institutions, which will conform with a) the new WHO model of antenatal care for basic routine care (10) and b) local protocols for special cases, standardized by us from previous trials in pre-eclampsia, hypertension,

urinary tract conditions, and intra- and post-partum care. An important conceptual issue is that we do not aim to detect any new, unexplored risk factors. Rather, we plan to determine how risk factors, that are routinely recorded during standard antenatal care, are distributed in the preterm/IUGR cases and their subgroups across these populations, in view of the considerable heterogeneity in risks and outcomes within these two conditions.

We plan to investigate the determinants of preterm delivery and IUGR in sub-groups from clinical, routine laboratory, demographic and socio-economic variables obtained from all women attending the study centres without any exclusion (as opposed to FGLS which aims to produce standards from a sample of selected, healthy women). It would, of course, be very interesting to collect more detailed information about other variables or test biomarkers of, for example, infection. The question, as always, is when to stop adding more variables to an already complex study. One alternative is to add ancillary studies to selected centres that are interested in collaborative research with other groups. It would certainly not be impossible for us to coordinate the collection, storage and testing of such samples (as our group has experience of collecting thousands of samples for similar studies) but funding is not available for such an activity at this point.

We will apply standard statistical strategies (e.g. logistic regression analysis) using the ~4,500 preterm newborns or IUGRs from the total NCSS study population to study these relationships compared to term babies, and we shall explore some recently proposed analytical strategies as well (35). We will consider subgroups of preterm deliveries, e.g. induced vs. spontaneous deliveries; premature rupture of membranes; pre-eclampsia related vs. unexplained impaired fetal growth, as well other pathological and physiological conditions (36, 37). The adjusted ORs and confidence intervals for these risk factors as related to the different subgroups will be estimated and biologically plausible interactions will be explored. The association between subgroups of newborns and perinatal outcomes will be also examined. We also intend to explore several factors that might explain variability in fetal growth, e.g. parity, BMI, gestational age, position/presentation of the fetus, liquor volume, gender and number of measurements. We shall also introduce a comprehensive system for evaluating intra- and inter-observer variability using the 2D and 3D measures, as well as a random sample of retaken measures by an external ultrasonographer in each centre.

D. Objective IV: Acquisition of additional 3D images to create an anatomical and growth databank of individual fetal organs (Activities 8 - 9)

Activity 8: 3D Data Collection during FGLS

Data collection for this activity will be conducted immediately after the 2D measurements have been taken for FGLS. The 3D volumetric images of the head, abdomen, and femur will be acquired using the methods described in Section 1.6. We expect that the provider of the equipment will provide technical assistance to optimize data collection.

The primary purpose of the 3D component is to provide volumetric data to assess the quality of the 2D measures (BPD, OFD, HC, TAD, APAD, AC and FL). The secondary objectives are a) to evaluate currently used ultrasound measures of fetal volume with the expectation that some of them can contribute to the improvement of the presently used estimation of fetal weight and b) to establish a unique "biological databank" for future research. Additional 3D volumes of the chest (at the level of the 4-chamber view of the heart) and the humerus will be taken specifically for these purposes in some centres.

The secondary objective a) is very interesting from a clinical point of view as most of the ultrasound-derived equations for estimation of fetal weight have low predictability and were developed using data collected from small studies with less sophisticated ultrasound equipment than is now available. Thus, the feasibility of combining 2D and 3D measures will be explored including the possibility of doing so across gestational ages. We expect to evaluate the additional benefits that could be obtained by adding 3D values in a sub-sample of the study.

Other uses of the 3D data are planned: for example, the accuracy of fetal weight estimation based on 2D measures is known to be poor in clinical practice. It is expected that 3D data will improve the formulae that are still in use despite being produced over 20 years ago. Furthermore, the likelihood of low-cost, portable 3D equipment being available in the future raises the possibility of managing high-risk pregnancies in rural or distant regions by sending the images to a referral centre rather than the mother herself. We plan to take the final decision regarding the amount of 3D data to be collected and the number of participating sites after conducting a pilot study assessing the human resources available at the centres, the need for additional training, the time that 3D scanning adds to each visit, equipment considerations and the feasibility of transferring the very large datasets to the Coordinating Unit in Oxford. Issues we shall explore in the pilot study include: a) difficulties obtaining individual organ volumes (38); b) the need to obtain at least two good volume measures from most of the areas of interest which may require 20-30 minutes additional scanning; c) the need for the audit staff to know the values of the calculated volumes and the way the operator arrived at each volume which could introduce considerable bias, and d) how to save and transmit the volumetric data reliably given the risks of a hard drive crashing or being corrupted. In any case, it is unlikely that we shall record more than the 5 volumes listed above at each visit, and it is also worth mentioning that the Gates Foundation has not funded the 3D component of the study.

We have debated the issue of adding more ultrasound measures (i.e. we have a unique opportunity that may not be repeated again) with the Ultrasound Advisory Group and other experts in the field, as we are concerned that the potential list of additional 2D/3D measures is long. As is the case in selecting which data to collect in any large study, one has to balance the enthusiasm of the researchers against pragmatism. We have chosen to concentrate on taking a few measures well (and more than once) in the time available during each scanning session rather than risk the possibility that additional measurements (including some not used in routine clinical practice) will be taken hurriedly. Of course, it will be possible to take further 2D measurements retrospectively for research purposes, if needed, from the acquired 3D volumes.

Activity 9: Preparation of 3D Files and Exploratory Analyses

There are two goals of the 3D component of this project. The first is to understand better the clinical advantages of making manual and automated fetal measurements on 3D v. 2D data. This should lead to a clearer understanding of IUGR and the relationship between different growth patterns and fetal/neonatal outcomes, as well as the impact of delayed growth in individual fetal organs and health care problems in adult life (e.g. an increased risk of osteoporosis). The second goal is to derive statistical models of the fetus or parts of the fetus to provide informative visualizations of nominal shapes and growth data at different gestational ages.

One possible approach will be to build parts model of the fetus, and to deform it to different instances of data acquired at the same gestational age. This will not be a simple task and will require theoretical and technical advances in biomedical image analysis. The research will be led by A.Noble, Professor of Engineering Science, University of Oxford, and forms the basis of a separate grant application to a UK agency, the UK Engineering and Physical Sciences Research Council. The aims of Activity 9 are simply therefore to prepare the 3D images and commence these exploratory analyses. Examples of what is currently possible include studying the relationship between a) brain growth and neonatal/infant neurodevelopment (39) and b) fetal bone growth and neonatal/infant bone density to assess the origins of osteoporosis (40).

Ethical Issues

The study will comply with the International Ethical Guidelines for Biomedical Research Involving Human Subjects. Ethical safeguards will include:
Ethical approval at international, national and local levels: The study protocol will be submitted to the relevant UK ethics committee in keeping with the University of Oxford's

© University of Oxford

responsibilities as the study sponsor, as well as the appropriate national bodies of the countries included in the study (if required) and all local ethics committees. Ultrasound and anthropometric research involves minimal, if any, risk to the mother and fetus and is part of routine care in these institutions.

Individual informed consent: Written informed consent will be obtained from all women enrolled in the study after the study objectives and procedures have been described in detail. The information will be provided in written form (translated into all the required languages) and orally by the researchers themselves based on a standard template, which can be referred to whenever needed. Women will only be recruited if it is clear they understand what the research entails and once they have had sufficient time (at least 48 hours) to consider whether to participate or not. All information collected in these studies will be made available to the clinicians responsible for the women's care; however, the 2D ultrasound measurements will only be made available after the data have been submitted on-line (see 1.2 above).

Discontinuation: Women who have agreed to participate in FGLS may decide to leave the study at any time without adversely affecting their care in any way.

Confidentiality: All phenotypic and imaging data will be available for clinical use at each centre. For research purposes, privacy rules will be maintained in line with the *Health Insurance Portability and Accountability Act of 1996* (HIPAA). All individuals will be de-identified to ensure confidentiality. At the Coordinating Unit, all phenotypic and imaging data will therefore be stored, transmitted and analyzed anonymously.

Data sharing for clinical care: All the clinical data we plan to collect are being obtained anyway as part of routine clinical practice in these institutions; the data will therefore will be available to the providers at all times. Our proposed policy with regard to the ultrasound measures is as follows: a) the gestational age estimate will be incorporated into the medical records; b) an ultrasound examination for structural congenital malformations will be performed at the time of the 3rd scan, if this conforms with local practice, and the results will be incorporated into the medical records; c) as far as 2D ultrasound measurements are concerned, after the blinded values have been submitted electronically to the dataset, the last measurement will be provided for clinical use, and d) no 3D data will be provided for clinical use.

Finally, there are concerns about the misuse of ultrasound for sex selection in places where it may be used (especially if the practice is illegal). We shall therefore be very vigilant in regions where this practice is known to occur and monitor sex ratios in the sample continuously.

E. Challenges

Implementing this project presents major challenges for the research team. However, our experience in conducting large-scale studies in pregnancy means that we are well prepared to face the challenges, some of which may be difficult to resolve:

Recruitment: We do not anticipate capacity problems in recruiting a large number of women (4,000) from 8 centres, but it remains unclear what proportion of women in each centre will actually be eligible for FGLS given the inclusion criteria. We anticipate an average eligibility of 75%, and expect the centres to devise local strategies to maximize that number.

Follow-up: We do not anticipate problems retaining pregnant women during follow-up but we do recognize that such a large follow-up study requires a commitment from participants and staff. In addition, the planning and coordination of the ultrasound scans with only one ultrasound machine available per centre will require considerable organization in the hospitals that should not be underestimated. Follow-up of the newborns presents the following challenges: a) difficulties measuring very small infants, especially those in an incubator; b) morbidities associated with prematurity, which may modify the measures to be taken; c) mortality affecting the final sample size, which explains why we intend focusing on healthy preterms, and d) follow-up after discharge is a problem in any such study, but we expect that adherence to the protocol will be very good for the preterm babies given the parents' concerns for their health.

Data quality: Standardization of a large number of health professionals is always a major challenge for which we are prepared. Coordinating travel arrangements between centres to monitor data quality will need to be well organized.

Data analysis: Constructing growth curves is a sophisticated task for which several statistical methods are available; the experience gained in the WHO study is a major asset to overcome this challenge.

Sample size: Selecting an adequately powered sample size for studies like this one is always a compromise between statisticians' estimates, logistics and cost. We are concerned that we should not over-extend our study by attempting to answer too many questions. Ideally, we should have a sample size to create preterm standards with stable outer centiles and birth weight for gestational age standards with sufficient newborns with severe outcomes across the gestational age distribution. We would also like to explore ethnic-specific growth, in the event that multiple ethnic differences emerge from the data. However, based on the findings of the WHO Child Growth Study, we consider it unlikely that each ethnic group will have its own distinct growth pattern. Nevertheless, we shall explore this question by including women from each of the 5 ethnic groups: Whites, Latin Americans, African blacks, Orientals and Indians.

Motivation: The final challenge, but perhaps the most important, is to maintain motivation of the local investigators and data collection teams throughout the project, as they are the people who can guarantee the quality of the study.

Resources: Finally, we appreciate that advocating the introduction of new fetal growth standards into clinical practice in developing countries could be criticized on the grounds that it will require already under resourced institutions to purchase relatively expensive ultrasound equipment. We accept the possibility that some institutions will purchase ultrasound equipment based on the results of this study, despite our recommendation that serial ultrasound measurements are not necessary for the evaluation of fetal growth during routine antenatal care (10). However, many district level and referral hospitals in developing countries already have ultrasound machines and when they are used to evaluate high risk pregnancies and/or women with uncertain gestational age, the results may be misinterpreted or misused as the charts available are often inadequate. Therefore, much of this project is orientated towards more effective use of equipment that is already in such institutions.

Interestingly, the expensive and very sophisticated 3D imaging system being investigated in this study, for the first time on this scale, has potential to benefit high-risk pregnancies in rural areas and isolated communities. The rationale is that it will become feasible to transfer images from peripheral clinics to specialized centres for reconstruction and interpretation as the cost of the equipment falls and it becomes more portable. Such a strategy would reduce the transfer of large numbers of women unnecessarily as occurs at present and contribute to the decentralization of antenatal care services worldwide.

F: Regional distribution of Study Sites:

We have recruited study sites in the following regions of the world: Europe (UK and Spain); Central and South America (Cuba and Brazil); Africa (Kenya); Arab World (Oman); Asia (China) and the India subcontinent (India).

G. Intervention Effectiveness for Impaired Fetal Growth

We ourselves have explored the question of what to do after IUGR is diagnosed and we have published a series of review papers critically evaluating RCT interventions for IUGR (41-44) and another group's recent update (45). There are not many interventions for treating (as opposed to preventing) IUGR, other than planned elective delivery, to which use of ultrasound contributes greatly in developing countries. This explains why referral to an adequate level of perinatal care is so important an option. Our study therefore focuses on helping clinicians to detect impaired fetal growth accurately, which should avoid incorrect diagnoses and, thereby, iatrogenic preterm births. The misuse of this technology is one of the factors responsible for unnecessary medical interventions and why some treatments or interventions are shown to be ineffective, i.e. treatment of fetuses that may not have

required it in the first place. Furthermore, in the light of information contained in the Lancet 2008 Nutrition Series and other similar recent reports, we will consider the possibility of implementing nested case-control studies to obtain information related to maternal body composition, metabolism and micronutrient status, or growth factors and birth outcomes. We will explore them in detail as this is a unique opportunity for such studies.

We think that interventions can only be effective if they focus on the factor responsible for the growth restriction. IUGR is a very heterogeneous condition, as we have shown, and it is unlikely that a "silver bullet" will resolve all IUGR (or preterm deliveries), even if correctly detected. Overall, we believe that accurate early identification, avoiding false positive cases, and appropriate referral to adequate levels of obstetric and newborn care will have a great impact on neonatal mortality and resource allocation in developing countries.

VIII. Monitoring, Evaluation, and Dissemination

A. Monitoring and Evaluation

Data collection will be monitored using routine procedures and protocols employed by our network in several similarly sized, multi-centre studies. The production and evaluation of specific standards will undergo a methodical and rigorous peer review process through the expert consultation mechanisms used in previous projects: the WHO Child Growth Study and the perinatal multi-centre studies conducted by our network. Monitoring the implementation process at the study sites will follow principles established in implementing previous, similarly complex, studies carried out by Consortium members. This will include site visits and on-going quality control measures (see Section IV).

B. Dissemination of Results

The generous participation of hundreds of scientists in all aspects of dissemination was the key to the success of the WHO Child Growth Study and we will therefore rely on their already established system. So, involving the same WHO team in disseminating the results locally will be our principal strategy; this includes local and regional meetings and symposia. We will then focus on the local and global professional societies that are an integral part of the application, as well as health authorities in the participating countries and regional health institutions. Finally, a formal dissemination strategy similar to the one used by the WHO Child Growth Study will be implemented, including partnerships with other key players in the international field such as UNICEF, other UN agencies, NGOs and government institutions.

IX. Optimizing Public Health Outcomes and Intellectual Property Plans to Achieve Global Access

It is expected that these international standards will be used to assess the growth of fetuses and newborns throughout the world. The basic assumption behind the proposed standards is that fetal growth is optimal in healthy populations. Therefore, the proposed curves will constitute "optimal" standards. Widespread uptake of these fetal curves will be ensured by ultrasound equipment manufacturers incorporating them into their systems. The potential uses are:

1) Population assessment to a) provide a reference for comparison of the means (or medians), standard deviations, and trajectories of the population means (or medians) for any given sample, b) estimate the proportions of fetuses-newborns below a given cut-off in a sample, and c) standardize for gestational age and sex differences between samples, thereby allowing comparisons of prevalence rates below a given cut-off across samples.

2) Individual assessment to a) screen for fetal growth on a single occasion to identify fetuses with excess or deficient growth leading to appropriate diagnostic and/or therapeutic interventions (including delivery), and b) provide a clinical tool for assessing the efficacy of treatments or helping to choose time of delivery, particularly in severely growth restricted fetuses. Hence, the primary use of the curves at an individual level will be to identify unacceptable deviations as early as possible to prevent severe deficits and excesses, although they will need to be introduced into clinical practice with appropriate care and auditing. We expect to link with, and build upon, WHO's global efforts to disseminate the child growth charts. The large network of proposed partners and consortium members should prove, with the help of industry, to be a very effective method for disseminating the results at both clinical and public health levels.

A. Intellectual Property (IP) Plan.

	Yes	No
1. Is the proposed research likely to lead to any patentable or commercially exploitable results?		X
The study will provide each centre with a sophisticated (commercially available) 3D ultrasound machine, specially adapted for the needs of the study, as well as technical expertise for data transfer and storage of 2D and 3D data. The proposed research is unlikely to lead to patentable results. Ultrasound machine manufacturers throughout the world will be encouraged as part of the Dissemination and Implementation Plan to incorporate the new standards into their machines from which they may derive commercial benefit.		

	Yes	No
2. Will the proposed project, either at its inception or at a foreseeable future point, depend on the use of technologies, materials, or other inventions that may conflict with goals of global access in terms of either cost or availability in the developing world?		X

	Yes	No
3. Is the proposed project and related IP subject to any agreements (e.g., licenses, collaborations, research or funding agreements or any other form of agreement) with commercial, academic, or other organizations, including other funding entities, subgrantees or subcontractors?		X

	Yes	No
4. Does your organization plan to assume responsibility for maturation, production, and dissemination of the innovation itself?	X	

© University of Oxford

B. Commitment to Sharing Data and Materials

It is highly unlikely that patentable IP will be generated by this project. The growth standard and related analyses represent the new evidence-based knowledge that will be generated. These will be widely disseminated, especially in open access publications, and shared with research and clinical communities worldwide, as well as the appropriate national and international agencies.

The Fetal Growth Standards charts will be made available to all manufacturers of ultrasound machines. Data from each centre will be provided to the local investigators with the understanding that the primary and secondary objectives of this study can only be explored with the pooled data set and can not be replicated using local data. Other locally relevant issues can be analyzed by individual centres.

The study Steering Committee (SC) and Executive Committee (EC) will also, after the main tasks are completed, engage in negotiations with other *bona fide* researchers for access to the data set to allow scientific and public health relevant questions to be explored. The SC and the EC have developed a set of rules and procedures for the evaluation of such requests in keeping with the commitment of the present investigators to ensure the widest possible utilization of the data (see Appendix F).

C. How will we bridge the new data to the existing data?

We aim to produce the following new standards:

1. **Fetal growth by ultrasound:** Based on discussions with ultrasound manufacturers and the known limitations of existing charts, we anticipate that the newly developed fetal growth charts will replace all charts that are currently incorporated into ultrasound machines.

2. **Postnatal growth for preterm babies:** Based on discussions with a number of pediatricians in different countries and WHO information, we anticipate that the postnatal growth charts for preterm babies will be incorporated into routine clinical practice without much difficulty because there is a lack of well-developed charts for this specific purpose. We believe that the international community is looking forward to such charts being developed.

3. **Birth weight for gestational age standards:** It is likely that the new, international, risk-related, birth weight for gestational age standards will be widely adopted following implementation of the new fetal growth and preterm postnatal growth standards. However, we acknowledge that this process may take longer because other standards are already being used (even though their limitations are recognized). We shall therefore work closely with WHO to follow the strategy they are using for the infant growth standards.

We do not anticipate any conceptual limitations to the future use of the standards, particularly in the light of the statement in the recently published Lancet "Maternal and Child Undernutrition" series that international fetal and newborn growth standards need to be developed (5). Nevertheless, it is clear that whenever efforts are being made to change clinical practice (even if the need to do so is recognized), common barriers and resistance to change will be encountered. We feel that our experience in introducing the MGRS standards will be invaluable for this task although, in theory at least, introducing the ultrasound standards should present fewer problems as they will be built into the equipment and we hope they will be endorsed by the International Society of Ultrasound in Obstetrics & Gynecology and other professional associations responsible for recommending new technology into clinical practice.

Rolling out the fundal-height charts and the newborn and preterm growth standards will use the same procedures as WHO did when it rolled out its Child Growth Standards. We shall coordinate these processes from the start of INTERGROWTH-21st by liaising closely with both WHO and the leading manufacturers of ultrasound machines. The WHO Child Growth Standards team has considerable experience in rolling out growth standards globally. Within only 18 months of starting to implement those standards, 72 countries have officially adopted them and 80 are in the process of doing so. The network already developed by the WHO team will serve as a means of rolling out the standards we propose

developing in INTERGROWTH-21st. In reality, many of the countries currently implementing the WHO Child Growth Standards have expressed a need for standards for fetal growth and newborn and preterm infants.

X. Organizational Capacity and Management Plan

A. Organizational Capacity and Facilities

History: The University of Oxford enjoys an international reputation as a world-class centre of excellence in research and teaching. It employs over 7,800 academic, research and support staff across a wide range of academic disciplines. The Medical Sciences Division, within which the Nuffield Department of Obstetrics & Gynaecology (NDOG) is located, is one of the major centres for clinical and basic biomedical research in Europe, with more than 2,200 staff and 800 postgraduate students. It achieved top scores in both the 1996 and 2001 UK HEFCE research assessment exercises, and fosters the highest possible standards in research, teaching and patient care. The annual grant income from external sources for the Medical Sciences Division is approximately $200M. The WHO has a long history of introducing programmes into countries and the on-going process of dissemination and implementation of the MGRS is a model that will be used as part of WHO's contribution to the project.

Experience: The Consortium members (University of Oxford, WHO Nutrition and collaborating centres) have extensive experience in conducting multi-centre studies and constitute perhaps the largest research network in the world in the field of women's and perinatal health. The WHO MGRS (8) and the multi-centre, epidemiological and clinical studies conducted by the Maternal & Perinatal Research Network are clear examples of the Consortium's organizational skills and capacity for interacting with regional/local health authorities, universities and leading institutions worldwide. NDOG has coordinated large-scale genetic epidemiology studies across multiple international sites, involving data collection from thousands of individuals in the ENDOGENE Study. Office space within its existing facilities at the John Radcliffe Hospital, Oxford, will be provided to house the Coordinating Unit (CU) staff, as well as the necessary IT support and administrative assistance. Sites have been chosen on the understanding that they have the necessary infrastructure to conduct the study (e.g. IT and capacity for ultrasound research). Finally, the professional societies recruited to help coordinate implementation and dissemination of the results are among the leading groups in the world in the field of obstetrics.

B. Management and Staffing Plan

Management: The study will be coordinated and managed by NDOG (University of Oxford), where the CU will be located (see Appendix F). Four units responsible for day-to-day monitoring of quality control and data collection (2D Ultrasound Data Quality Unit; Anthropometric Data Quality Control Unit; Data Management Unit and 3D Ultrasound Data Quality Unit) will report directly to the CU. All data, except the 3D data, will be centralized at the Data Management Unit.

The Steering Committee (SC) consists of representatives of Consortium members and the lead investigator from each of the study centres. It has also external members to provide support in areas related to the study. The SC will meet twice a year to review progress, ensure uniformity of data collection from the study sites, and discuss any substantive issues that arise. Any adaptations to the protocol or technical variations required at individual sites because of local needs will be reviewed and approved by the SC.

The Executive Committee (EC) will monitor the progress of the study on a regular basis and resolve substantive issues arising from implementation of the study. The Study Coordinator will be ex-officio member of the EC. The EC will make the final selection of study sites; exclude a site if data collection standards are not sufficiently high, and make final decisions about sample selection and analytical issues related to construction of the standards.

The AC, consisting of internationally recognized experts in ultrasound, obstetrics, statistics, nutrition, fetal development, anthropometrics, epidemiology and biology, will

provide technical advice to the CU, EC and SC. The study will be indemnified by the University of Oxford that will carry liability for any adverse outcomes arising from breaches of research protocol.

Study coordination:

The **Principal Investigator** (José Villar) will be responsible for all scientific and technical components of the project. The **Project Director** (Stephen Kennedy) will be the focal person in NDOG responsible for overall study management. The **Project Leader** (Layla Sheikh), will be responsible for day-to-day implementation of the study, coordinating and monitoring of all activities as well as training, monitoring and evaluation. A **Statistical Director** (Doug Altman) will advise on sampling strategies, sample sizes and all statistical methods, in particular those for the construction of the new standards. He will be assisted in Years 1-5 by a **Data Management Supervisor** (Lucy Hoch), who will coordinate all aspects of data management between the Data Management Unit (DMU), the CU in Oxford and the other data quality units. This will include the preparation of data sets, quality control strategies and data files for use by the team of statisticians. This strategy for data management and analysis has been extremely efficient in previous, large multi-centre studies conducted by us as it allows for direct contact between the data collection centres and the coordinating unit. In Years 3-5, Dr Altman will also be assisted by a **Senior Statistician** to develop, test and implement computer programs to construct the new growth standards.

There will be a **Senior Technical Coordinator** (Aris Papageorghiou, St George's, London) of the ultrasound component of the study, who will have responsibility for overall coordination of the ultrasound measurements, including quality control and standardization. He will be assisted by an **External Ultrasound Expert** (TBN), acting as a consultant and a **Clinical Research Fellow** (TBN) to oversee the quality of a random sample of ultrasound measurements during the data collection period. The CU team will be completed by a **Secretary** (Melissa Shorten) responsible for secretarial support to the CU, a **Project Administrator** (Ann Lambert) in charge of all aspects of transferring funds to the study centres, equipment, meetings, travel and coordination between the CU, DMU and study centres, and a **Research Associate** (Hannah Knight) responsible for piloting protocol and monitoring its implementation. Finally, **WHO Focal Point** (Mercedes de Onis) will be a member of the SC, EC and the CU and will liaise between the different units of the study and WHO, so as to integrate the previous study fully into the present project.

XI. Citations

1. Lawn JE, Cousens S, Zupan J. 4 million neonatal deaths: when? Where? Why? Lancet 2005;365:891-900.

2. Williams RL, Creasy RK, Cunningham GC, Hawes WE, Norris FD, Tashiro M. Fetal growth and perinatal viability in California. Obstet Gynecol 1982;59:624-32.

3. Rao SC, Tompkins J. Growth curves for preterm infants. Early Hum Dev 2007;83:643-51.

4. de Onis M, Habicht JP. Anthropometric reference data for international use: recommendations from a World Health Organization Expert Committee. Am J Clin Nutr 1996;64:650-8.

5. Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M et al. Maternal and child undernutrition: global and regional exposures and health consequences. Lancet 2008;371:243-60.

6. de Onis M, Onyango AW, Van den Broeck J, Chumlea WC, Martorell R. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. Food Nutr Bull 2004;25:S27-36.

7. Kramer MS, Platt RW, Wen SW, Joseph KS, Allen A, Abrahamowicz M et al. A new and improved population-based Canadian reference for birth weight for gestational age. Pediatrics 2001;108:E35.

8. de Onis M, Garza C, Onyango AW, Martorell R. WHO Child Growth Standards. Acta Paediatr 2006;450:1-101.

9. Zhang X, Platt RW, Cnattingius S, Joseph KS, Kramer MS. The use of customised versus population-based birthweight standards in predicting perinatal mortality. BJOG 2007;114:474-7.

10. Villar J, Ba'aqeel H, Piaggio G, Lumbiganon P, Miguel Belizan J, Farnot U et al. WHO antenatal care randomised trial for the evaluation of a new model of routine antenatal care. Lancet 2001;357:1551-64.

11. Kelly Y, Panico L, Bartley M, Marmot M, Nazroo J, Sacker A. Why does birthweight vary among ethnic groups in the UK? Findings from the Millennium Cohort Study. J Public Health 2008;epub Jul 21.

12. Villar J, Valladares E, Wojdyla D, Zavaleta N, Carroli G, Velazco A et al. Caesarean delivery rates and pregnancy outcomes: the 2005 WHO global survey on maternal and perinatal health in Latin America. Lancet 2006;367:1819-29.

13. Villar J, Abdel-Aleem H, Merialdi M, Mathai M, Ali MM, Zavaleta N et al. World Health Organization randomized trial of calcium supplementation among low calcium intake pregnant women. Am J Obstet Gynecol 2006;194:639-49.

14. Gulmezoglu AM, Villar J, Ngoc NT, Piaggio G, Carroli G, Adetoro L et al. WHO multicentre randomised trial of misoprostol in the management of the third stage of labour. Lancet 2001;358:689-95.

15. Royston P. Constructing time-specific reference ranges. Stat Med 1991;10:675-90.

16. Kramer M, Victora CG. Low birth weight and perinatal mortality. In: Nutrition and Health in Developing Countries. Totowa, NJ: Humana Press, 2001:57-69.

17. Carroli G, Rooney C, Villar J. How effective is antenatal care in preventing maternal mortality and serious morbidity? An overview of the evidence. Paediatr Perinat Epidemiol 2001;15 Suppl 1:1-42.

18. Villar J, Carroli G, Zavaleta N, Donner A, Wojdyla D, Faundes A et al. Maternal and neonatal individual risks and benefits associated with caesarean delivery: multicentre prospective study. BMJ 2007;335:1025.

19. Merchant KM, Villar J. Maternal height and newborn size relative to risk of intrapartum caesarean delivery and perinatal distress. BJOG 2001;108:689-96

20. Goedhart G, van Eijsden M, van der Wal M, Bonsel G. Ethnic differences in term birthweight: the role of constitutional and environmental factors. Paediat Perinat Epidemiol 22:4:360-368

21.   Mortensen LH, Diderichsen F, Arntzen A, Gissler M, Cnattingius S, Schnor O et al. Social inequality in fetal growth: a comparative study of Denmark, Finland, Norway and Sweden in the period 1981-2000. J Epidemiol Community Health 2008;62:325-31.

22.   Lynch CD, Zhang J. The research implications of the selection of a gestational age estimation method. Paediatr Perinat Epidemiol 2007;21 Suppl 2:86-96.

23.   Robinson HP, Fleming JE. A critical evaluation of sonar "crown-rump length" measurements. Br J Obstet Gynaecol 1975;82:702-10.

24.   Group WMGRS, de Onis M. Complementary feeding in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 2006;450:27-37.

25.   Neufeld LM, Haas JD, Grajeda R, Martorell R. Ultrasound measurement of fetal size in rural Guatemala. Int J Gynaecol Obstet 2004;84:220-8.

26.   Blake KV, Gurrin LC, Beilin LJ, Stanley FJ, Kendall GE, Landau LI et al. Prenatal ultrasound biometry related to subsequent blood pressure in childhood. J Epidemiol Community Health 2002;56:713-8.

27.   British Ultrasoud Society. Fetal Measurements Document, 2008.

28.   Verburg BO, Steegers EAP, De Ridder M, Snijders RJM, Smith E, Hofman A, Moll, HA, Jaddoe VWV, Wittema JCM. New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. Ultrasound Obstet Gynae:31:388

29.   Thorsell M, Kaijser M, Almstrom H, Andolf E. Expected day of delivery from ultrasound dating versus last menstrual period--obstetric outcome when dates mismatch. BJOG 2008;115:585-9.

30.   Platt LD. Three-dimensional ultrasound, 2000. Ultrasound Obstet Gynecol 2000;16:295-8.

31.   Belizan JM, Villar J, Nardin JC, Malamud J, De Vicurna LS. Diagnosis of intrauterine growth retardation by a simple clinical method: measurement of uterine height. Am J Obstet Gynecol 1978;131:643-6.

32.   McGuire W, Henderson G, Fowlie PW. Feeding the preterm infant. BMJ 2004;329:1227-

33.   Villar J, Abalos E, Carroli G, Giordano D, Wojdyla D, Piaggio G et al. Heterogeneity of perinatal outcomes in the preterm delivery syndrome. Obstet Gynecol 2004;104:78-87.

34.   Borghi E, de Onis M, Garza C, Van den Broeck J, Frongillo EA, Grummer-Strawn L et al. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. Stat Med 2006;25:247-65.

35.   Bhandari N, Bahl R, Taneja S, de Onis M, Bhan MK. Growth performance of affluent Indian children is similar to that in developed countries. Bull World Health Organ 2002;80:189-95.

36.   Owusu WB, Lartey A, de Onis M, Onyango AW, Frongillo EA. Factors associated with unconstrained growth among affluent Ghanaian children. Acta Paediatr 2004;93:1115-9.

37.   Kramer MS, Ananth CV, Platt RW, Joseph KS. US Black vs White disparities in foetal growth: physiological or pathological? Int J Epidemiol 2006;35:1187-95.

38.   Zhang J, Villar J, Sun W, Merialdi M, Abdel-Aleem H, Mathai M et al. Blood pressure dynamics during pregnancy and spontaneous preterm birth. Am J Obstet Gynecol 2007;197:162 e1-6.

39.   Villar J, Carroli G, Wojdyla D, Abalos E, Giordano D, Ba'aqeel H et al. Preeclampsia, gestational hypertension and intrauterine growth restriction, related or independent conditions? Am J Obstet Gynecol 2006;194:921-31.

40.   Chang CH, Yu CH, Ko HC, Chen CL, Chang FM. Predicting fetal growth restriction with liver volume by three-dimensional ultrasound: efficacy evaluation. Ultrasound Med Biol 2006;32:13-7.

41.   Huang H, Zhang J, Wakana S, Zhang W, Ren T, Richards LJ et al. White and gray matter development in human fetal, newborn and pediatric brains. Neuroimage 2006;33:27-38.

42.   Cooper C. Epidemiology of osteoporotic fracture: looking to the future. Rheumatology (Oxford) 2005;44 Suppl 4:iv36-iv40.

43.   Villar J, Gulmezoglu AM, de Onis M. Nutritional and antimicrobial interventions to prevent preterm birth: an overview of randomized controlled trials. Obstet Gynecol Surv 1998;53:575-85.

44.   Gulmezoglu M, de Onis M, Villar J. Effectiveness of interventions to prevent or treat impaired fetal growth. Obstet Gynecol Surv 1997;52:139-49.

45.   de Onis M, Villar J, Gulmezoglu M. Nutritional interventions to prevent intrauterine growth retardation: evidence from randomized controlled trials. Eur J Clin Nutr 1998;52 Suppl 1:S83-93.

46.   Kulier R, de Onis M, Gulmezoglu AM, Villar J. Nutritional interventions for the prevention of maternal morbidity. Int J Gynaecol Obstet 1998;63:231-46.

47.   Iams JD, Romero R, Culhane JF, Goldenberg RL. Primary, secondary, and tertiary interventions to reduce the morbidity and mortality of preterm birth. Lancet 2008;371:164-75.

XII. Appendices

Appendix A: Data collection Forms

Three different types of data recording forms will be used:

FGLS:

1) evaluation for study entry form

2) ultrasound dating form

3) study entry form corresponding to the first antenatal care visit

4) follow-up visit forms for ultrasound and clinical data

5) pregnancy summary and delivery form

6) hospital admission and referral form

PPFS:

1) screening form for subject eligibility and newborn anthropometric form

2) a study entry form corresponding to either the first postnatal care visit after newborn hospital discharge or at 15 days post delivery if the preterm baby is still in the hospital

3) follow-up clinic and anthropometric visit forms

4) newborn hospital admission and referral form

NCSS (same two forms as corresponding ones in FGLS):

1) pregnancy summary and delivery form (including newborn anthropometrics)

The reasons for discontinuing the study will be recorded in a common **Discontinuation Form** to be used in both longitudinal studies. It should be emphasized that all forms are as short as possible to improve compliance. Therefore, all questions have been carefully evaluated to ensure that they serve at least one of the following purposes: a) eligibility (e.g. socio-economic status); b) sample description (demographic and environmental variables, etc.); c) exclusion criteria (major illnesses, etc.); d) standardization of results across centres; e) future use of references (vitamin/mineral supplements), or f) to assess possible selection biases.

Appendix B: DATA FLOW MODELS

FETAL GROWTH LONGITUDINAL STUDY DATA FLOW

Initial interview with potentially eligible patients

Complete **Screening Form** for new patients (clinical and medical history)

Does the patient meet the basic clinical/medical history entry criteria? (see p. 18, protocol)

YES

NO

Arrange ultrasound dating scan

Complete **Ultrasound Dating Form**

Is the difference between gestational age estimated by Crown Rump Length (CRL) and LMP less than 7 days?

NO

YES

Ask patient to give informed consent, complete **Patient Consent Form**

IF CONSENT NOT GIVEN

Send anonymized data to Data QC and Entry Unit

YES

Send **Consent Form** to hospital project coordinator to add to ethics file

Enter patient into study, provide subject study number and complete **Entry Form** (first study antenatal visit)

Send the **Entry Form** to Data QC and Entry Unit

Schedule first follow-up appointment for 5±1 weeks

© University of Oxford

FETAL GROWTH LONGITUDINAL STUDY DATA FLOW (CONT.)

Follow-up scan 1 → Complete **Pregnancy Follow-up Form 1**

Follow-up scan 2 → Complete **Pregnancy Follow-up Form 2**

...Follow-ups 3...4...5... → Complete **Pregnancy Follow-up Form 3...4...5...**

Follow-up scan 6 → Complete **Pregnancy Follow-up Form 6**

At any time if the woman is referred to a hospital or any other level of care...

...Please complete a **Referral Form**

Please send each **Pregnancy Follow-up Form** to Data QC and Entry

Please send any **Referral Forms** to Data QC and Entry Unit

At delivery, regardless of gestational age, complete **Pregnancy Summary and Delivery Form** and **End Report Form**

Please send **Pregnancy Summary and Delivery Form** and **End Report Form** to Data QC and Entry Unit

IS THE BIRTH PRETERM? (less than 37^{+0} weeks) — NO → The baby will complete the study at his/her hospital discharge. No more follow-up contact needed.

YES ↓

Complete **PPFS Screening Form** and enter the baby into PPFS. Take measurements if baby is stable

Please send **PPFS Screening Form** to Data QC and Entry Unit

After 3 days, is the preterm baby clinically stable and able to sustain life independently? — NO → The infant is not eligible for the study

YES ↓

Take measurements and Complete **Study Entry Form**

Send the **Study Entry Form** to the Data QC and Entry Unit

If the baby dies, complete **End of Study** Form at the time of death

Send the **End of Study Form** to the Data QC and Entry Unit

48

© University of Oxford

PRETERM DATA FLOW

Schedule first follow-up visit for 2 weeks time	Complete **Preterm Follow-up Form 1**	Send **Preterm Follow-up Form 1** to the Data QC and Entry Unit
Was the infant referred for another level of care or re-admitted to hospital? **YES**	Please complete **Hospital Admission and Referral Form**	Send the **Hospital Admission and Referral Form** to the Data QC and Entry Unit
NO ↓ Schedule follow-up visits every 2 weeks for 8 weeks, then every 4 weeks until 8 postnatal months	Complete **Preterm Follow-up Forms 2..3..4.......11**	Send all **Preterm Follow-up Forms** to the Data QC and Entry Unit
After visit 11 or death, the baby has completed the study		
Complete **End of Study Form**		Send **End of Study Form** to the Data QC and Entry Unit

© University of Oxford

Appendix C:

ULTRASOUND METHODS AND QUALITY CONTROL

Acknowledgements: The first version of the ultrasound protocol was prepared between 18th June and 10th July 2008 by Aris Papageorghiou and Laurent Salomon, with valuable input from Jan van den Broeck. Aris Papageorghiou revised the protocol based on feedback from the INTERGROWTH-21st Executive Committee and Paul Chamberlain. Further amendments were then made based on a series of exchanges between Aris Papageorghiou and Laurent Salomon, Jan van den Broeck (who also provided much of the material on quality control), Paul Chamberlain, José Villar and Stephen Kennedy. The protocol was finalized after extensive discussion at the initial meeting of the Steering Committee, which took place in Oxford on 15-17th September 2008.

1. Background

A central plank of INTERGROWTH-21st is the creation of fetal growth charts during the FGLS part of the study. The correct use of ultrasound is essential to ensure that accurate, reproducible and applicable results are obtained. This document sets out all the ultrasound requirements for the study, including measurements, training, equipment and quality control.

2.1 Initial ultrasound examination and dating:

All mothers will have a trans-abdominal ultrasound scan between 9^{+0} to 14^{+0} weeks. This will be performed to rule out ectopic pregnancy, missed abortion and multiple gestations and to estimate gestational age (GA).

For the purposes of the study, gestational age will be calculated by LMP and confirmed by fetal crown-rump length (CRL). In order to allow accurate assessment of gestation the inclusion criteria used were:
certain LMP
- Regular 26-30 day menstrual cycles in the last 3 months;
- No hormonal contraception use, pregnancy or breastfeeding in the 3 months preceding the LMP.
- As embryonic growth may be different in pregnancies from assisted reproductive techniques[2,3] only those with spontaneous conception were included
- The CRL will be plotted on well-established charts [4,5]. If the difference in gestational age estimation by CRL and LMP is ≤ 7 days we will consider the LMP to be reliable and take it as the true biological date. Women where the difference in gestational age between LMP and CRL is greater than 7 days will not be eligible.

2.2 Subsequent ultrasound examinations

Following the initial scan, women will have up to six scheduled scans at ~5 weekly (±1 week) intervals until delivery. When rescanning does not occur within the allotted time, the women will be asked to attend at their next scheduled study appointment.

Essential documentation

At all examinations after dating the following will be documented:
1. Fetal presentation (cephalic, Breech, Transverse, Oblique)
2. Placental localization (fundal, high anterior, high posterior, high right lateral, high left lateral, low anterior, low posterior, low right lateral, low left lateral).
3. Amniotic fluid volume (polyhydramnios, increased, normal, reduced, oligohydramnios, anhydramnios).

Essential measurements** (see **3. Methodology of ultrasound and definitions)*
At all examinations after dating the following measurements will be taken:
1. Biparietal diameter (BPD)
2. Occipito-Frontal Diameter (OFD)
3. Head circumference (HC) using the ellipse facility
4. Transverse abdominal diameter (TAD)
5. Antero-posterior abdominal diameter (APAD)
6. Abdominal circumference (AC) using the ellipse facility
7. Femur length (FL)

Optional measurements
This study presents a unique opportunity to obtain other measurements of fetal biometry. However, the introduction of extra tasks could reduce the overall quality of the seven main measurements and inconvenience the participants. Therefore, the feasibility of taking additional measurements will be evaluated in consultation with each centre before a final decision is made and only some centres are expected to participate in these additional studies. In descending order of importance:
1. Humerus
2. Radius / Ulna
3. Tibia / Fibula

2.5 3-Dimensional Ultrasound Volumes (3D ultrasound)

Essential 3D measurements
The objective of the 3D component is to provide volumetric data to assess the quality of the 2D measurements (BPD, OFD, HC, AC, FL). Hence, the 3D volumes that we obtain will be of the:
(1) head (capture 3D volume with 2D view at the level of the BPD)
(2) abdomen (volume at the level of the AC)
(3) femur (volume in the same view as measuring the FL)

Optional 3D measurements
In order to establish a unique "biological databank" for future research additional 3D volumes of other organs, e.g. the heart at the level of the 4-chamber view of the heart; placental volumes, can be taken specifically for these purposes. These will be considered ancillary studies and will be dependent upon local capacity and the approval of the Steering Committee.

We plan to take the final decision regarding the amount of 3D data to be collected and the number of participating sites for these ancillary studies after conducting a pilot study assessing the human resources available at the centres, the need for additional training, time that 3D scanning adds to each visit, equipment considerations and the feasibility of transferring the very large datasets to the Coordinating Unit in Oxford.

Issues that we will explore in the pilot study include:
a) difficulties obtaining liver, brain and fetal cerebellar volumes
b) the need to obtain at least two good volume measures from the areas of interest, which may require 20-30 mins additional scanning
c) the temptation to divert attention away from standard measurements to more "exciting new technology" which will be popular with women. For these purposes the surface rendering function will be disabled.
d) how to save and transmit the volumetric data reliably given the risks of a hard drive crashing or being corrupted.
e) With advancing gestation (especially in the third trimester, it is likely that we will not be able to record many of the 5 volumes listed above accurately (i.e. brain and abdomen).

© University of Oxford

3. Methodology of ultrasound and definitions

The first visit (dating scan) will be between 9^{+0} and 14^{+0} completed weeks. The aim of this scan is to confirm fetal viability, and to exclude multiple pregnancy or major fetal abnormality. Fetal CRL is measured at this visit. In women who meet the inclusion criteria and where the discrepancy between gestational age by LMP and CRL is ≤7 days, subsequent visits (for fetal biometry) will be scheduled at approximately 5 weekly (± 1 week) intervals (i.e. 14-18, 19-23, 24-28, 29-33, and 34-38 weeks and 39-42 weeks).

A full morphological evaluation will be conducted at 19-23 weeks following standard practice at each centre. Fetuses diagnosed with any minor abnormalities will be managed as per local clinical guidelines. If the clinical decision is to continue with the pregnancy the mother will remain in the study. Study follow-up for these cases will be identical, but eventual exclusion at the time of data analysis will be carried out.

Fetuses with major abnormalities that may affect morphometric measurements will be excluded from further study. All infants will receive anthropometrical assessment after delivery

3.1 Basic characteristics
1. Transabdominal ultrasound
2. Lateral recumbent position.
3. Essential ultrasound measurements are obtained at all visits
    - Biparietal diameter (BPD)
    - Occipito-Frontal Diameter (OFD)
    - Calculated Head circumference (HC) using the ellipse facility
    - Transverse abdominal diameter (TAD)
    - Anterio-posterior abdominal diameter (APAD)
    - Calculated abdominal circumference (AC) using the ellipse facility
    - Femur length (FL)
4. Measurements are obtained 3 times from 3 separately generated ultrasound images of each structure in a blinded fashion (no measurement visible) and submitted electronically.
5. Following this the final set of measurements are revealed for clinical management purposes as per local protocols.
6. In each view a single 3D volume is stored for quality control purposes.
7. Ultrasound equipment provides measurements to tenth of a millimetre.
8. Data are submitted electronically to the web-based data set
9. Depending on the availability of appropriate technology, associated images are also submitted electronically to the Coordinating Unit. If real time submission of images is not available these will be down-loaded onto a memory card and submitted via email.
10. Ultrasound images of CRL, BPD, OFD, HC, AC and FL must fill at least 30% of the monitor screen.

3.2 Measurement techniques for compulsory measurements [4-10]

Crown – Rump Length (CRL) – only at the dating scan (9^{+0} to 14^{+0} weeks)
Obtaining the image
- A mid-sagittal section of the fetus should be obtained
- The fetus should be horizontal (at 90° to the angle of insonation)
- The fetus should be in a neutral position (not hyperextended or flexed)
- *Magnification*
- The image must fill at least 30% of the monitor screen.

Caliper placement
- The intersection of the calipers should be placed on the outer borders of the head and rump.

Biparietal Diameter
Obtaining the image
- A cross-sectional view of the fetal head at the level of the thalami
- As close as possible to the horizontal with the angle of insonation as close as possible to 90°
- Oval shape
- Symmetrical
- Centrally positioned, continuous midline echo (falx cerebri) broken anteriorly at one third of its length by the cavum septum pellucidum
- The thalami should be located symmetrically on each side of the midline falx.

Magnification
- The cross section of the fetal head must fill at least 30% of the monitor.

Caliper placement
- The intersection of the calipers should be placed on the outer border of the parietal bones ('outer to outer') at the widest part of the skull.

Occipito-frontal diameter (OFD) and Head Circumference (HC)
Obtaining the image and magnification: obtained from the same still image as the BPD.

Caliper placement
The intersection of the calipers should be placed on the outer border of the occipital and frontal edges of the skull at the point of the midline ('outer to outer') across the longest part of the skull. The HC will be calculated from the BPD and OFD measurements using the ellipse facility (using the formula HC = π(BPD + OFD)/2).

Abdominal Circumference
Obtaining the image
- Transverse section of the fetal abdomen as close as possible to circular
- Umbilical vein in its anterior third
- Stomach bubble visible
- Kidneys and bladder not visible

Magnification
- The cross section of the fetal abdomen must fill at least 30% of the monitor screen.

Caliper placement
- The anterior-posterior abdominal diameter (APAD) and transverse abdominal diameters (TAD) are measured.
- To measure the APAD the intersection of the calipers is placed on the outer borders of the body outline from the posterior aspect (skin covering the spine) to the anterior abdominal wall.
- To measure the TAD the intersection of the calipers is placed on the outer borders of the body outline at 90° to the APAD, across the abdomen at the widest point.
- The AC will be calculated from the APAD and TAD measurements using the ellipse facility (based on the formula AC = π(APAD + TAD)/2).

Femur Length Technique
Obtaining the image
- To be imaged as close as possible to the horizontal plane.

© University of Oxford

- Angle of insonation of the ultrasound beam is 90°.
- The full length of the bone is visualised
- Not obscured by shadowing from adjacent bony parts.

Magnification
- The cross section of the fetal abdomen must fill at least 30% of the monitor.

Caliper placement
- The intersection of the calipers is placed on the outer borders of the edges of the femoral bone 'outer to outer').
- The trochanter is not to be measured.

3.3 Techniques for optional measurements:

The humerus, radius, ulna, tibia and fibula can all be measured from 14^{+0} weeks onwards. Similar to the guidelines of femur length measurement, the longest length of each bone is measured with the bone at as close as possible to 90° to the ultrasound beam. The humerus is measured from upper to lower lateral margins. In the forearm the ulna is distinguished from the radius by its longer proximal length (both radius and ulna end at the same level distally). In the lower leg the lateral bone is the fibula with the tibia lying medially to it. All long bones are to be measured along their longest lengths from upper to lower lateral margins. The bone measured should fill at least 30% of the width of the screen. **Only one side (the anterior or most easily accessible limb) should be measured.**

3.4 Inability to take measurements

The fetal position may on occasion be so persistently unfavourable as not to allow measurements according to the guidelines above.

CRL: Accurate measurement of CRL is of obvious importance in ensuring that a woman is eligible to take part in the study. If despite repeated attempts it is not possible to obtain a CRL measurement that allows confirmation of gestational age the woman is not eligible for the study (in the same way she will not be eligible for the study if the gestation calculated from CRL and LMP is discrepant by more than 7 days).

Essential measurements: With the exception of CRL, every effort should be made to obtain the best possible measurements taking into account the definitions above. This may require allowing the woman to go and return for the scan at a later date/time (within one week). Measurements should not be taken if it remains impossible to obtain a good quality image to allow accurate measurement this should not be taken. The next appointment should be kept as scheduled.

Optional measurements: The same principle applies to the optional measurements: these should not be taken if it is impossible to obtain a good quality image.

4. Equipment (based on RCR recommendations)[13]:

In all applications of ultrasound, three things are of prime importance: image clarity, resolution and the ability to differentiate tissue structures. Hence, although it is accepted that the best possible images may not be obtained from all patients at all times, the equipment chosen must be capable of visualizing tissue structures in the overwhelming majority of patients with different body sizes and shapes to a level sufficient to meet the exacting needs of the study.

The specifications below are the minimum requirements for the study, but it is important to recognize that image clarity and resolution may differ between manufacturers despite machines having similar specifications. It is therefore essential that candidate machines are tested before a final choice is made to assess their quality and to determine operator

© University of Oxford

preferences. It is envisaged that this will be done at the Study Coordinating Unit (Oxford) or the Ultrasound Coordinating Unit (St George¹s, London).

We realize that there is a potential trade-off between quality and cost and, therefore, both these factors need to be considered in a balanced way in making a decision.

Study specific considerations

1. All ultrasound equipment to be used will be standardized
   - Uniform probes
   - Uniform factory presets ("study preset")
2. Commercially available high quality real-time ultrasound scanner.
3. Less than 2 years old.
4. Transabdominal probes suitable for scanning throughout pregnancy.
5. Facility for on-line transfer of measurements and associated images.
6. Facility to "blind" measurements from examiner until after data transfer.
7. Facility to "unblind" the final measurement to allow clinical use.
8. Equipment will be serviced periodically by specialized technicians as per the manufacturer's recommendations.

Operation/ Display

General requirements:
- Operating from a single phase standard domestic 240V, 13A, 50 Hz mains supply (110V in Cuba)
- Large single high resolution non-interlaced monitor (min.15")
- Display examination details
    Basic patient ID
    date & time
    examination centre
    selected probe
    acoustic power setting
    mechanical and thermal indices and other image processing and Doppler information to be shown in all display modes on all video/disk/hardcopy outputs
- Floating keyboard and monitor with easy to move console
- Single and dual display of the modes should be available simultaneously
- High definition variable size display magnification
- Digital Processing Channels Beamformer
- Digital display memory
- Large Dynamic Range
- Automated 'Tissue Specific' pre-sets
- Automated 'Tissue Specific' signal processing
- Ability to store reasonable number of images
    The number of images that can be stored must be stated.
- Large Cineloop
- Extended signal processing facilities
- Customizable pre-sets and calculations, for individual users and for different types of applications, for all modes
- Automated 'INTERGROWTH-21st Study" pre-sets
- Full DICOM-3 image and Cineloop transfer activated
- Hard Disk storage
- CD rewriter or DVD rewriter for backup of presets, configuration and archiving of studies in DICOM format
- DVD ±R(W) Drive
- > 100GB on board archive
- S video or composite video out (ideally all of the following: RGB, composite, S video, VGA and DVI)

- Ability to save images / cineloops in jpeg / avi format to external media for teaching purposes and ideally the ability to store images / cine loops to USB key
- On screen measurement of distance, area and circumference (using ellipse or tracing methods), where a minimum of four simultaneous distance measurements or one area/ circumference measurement is required

- High definition B-mode
  - Controls of transmit power, receive gain (including TGC) and dynamic compression
  - Dynamic transmit and receive focusing with user selection of focal zones
  - Configurable spatial/temporal averaging
  - Measurement tools such as distance, area, circumference and 2D-area (ellipse or tracing)

Scanheads
- The scanner should have, at least 2 active transducer ports from which one can be selected from the control panel.
- Broad band 2D probe which must be suitable for examinations at 9-14 weeks, 21-23 week and third trimester growth scans and with a penetration suitable for a wide variety of patients.
  E.g. Broadband 3-6 MHz and broadband 5-9 MHz probes or a single broadband probe of 3-9 MHz

Accessories
- Thermal printer capable of printing onto continuous standard thermal paper
- DVD player/recorder

General Configuration
- The unit should be compact and provide reasonable portability.

Safety
- The unit should meet or exceed performance and safety requirements of UL 544, CSA C22.2 and IEC 60601-1 and be CE marked

Maintenance
- In order to ensure the equipment is maintained in proper working order, the basic maintenance requirements are listed below. Suppliers are required to specify the services they will provide and their costs:
  - Clear instructions on the regular maintenance that should be carried out by the user
  - If the Unit needs regular maintenance, the frequency with which any parts need replacing
  - Engineer call-out response time of 48 hours following initial assessment/advice from the hospital's maintenance team
  - Options available if the equipment is out of action for an extended period
  - Details of any maintenance contract(s) available to cover all normal device deterioration and emergency repairs
- The maintenance support should be such that there is automatic supply upgrade of new firmware

OPTIONAL **3D / 4D capability**
- 3D Broadband 3-6 MHz or 3D broadband 5-9 MHz probes
- Spatial Temporal Image Correlation (STIC) – option – capture full fetal heart cycle in real time using volumetric transducer
- Surface rendering **disabled**
- Volume Contrast Imaging
- Specific software for post-processing stored 3D volumes.

HD Flow imaging (bidirectional Power Doppler), Tissue Doppler, Spectral Doppler, Colour velocity mapping and Power Doppler mapping
- Controls of power, gain, velocity range, baseline shift, low pass filter and compression
- Control for display time base
- Manual and automatic spectral analysis and waveform index estimator
- Volume flow measurement, User defined calculations
- Good sensitivity, temporal and velocity resolutions
- Acceptable Range gate registration accuracy, Gate duration, Beam width, Penetration depth, Velocity, Direction indication, Directional discrimination, Velocity estimation accuracy, High-pass filter, Waveform index estimation accuracy, Volume flow estimation accuracy
- Acceptable Colour direction indication, Lowest detectable velocity, Highest detectable velocity, Image spatial resolution (axial, lateral and slice thickness), Temporal resolution, Velocity resolution, Tissue colour suppression, Angle dependence of colour, Registration colour/B-scan image, Penetration depth

Further Notes
- Any non-compliance with the specification must be clearly identified together with any alternative or additional features.
- A decision may be taken to lease the selected device. If so, a tendering exercise will be undertaken to choose an appropriate leasing company.
- The availability of any training including courses offered / recommended by the manufacturer / supplier should be stated, together with any associated costs that may be incurred. Preference will be given to manufacturers / suppliers offering free fault finding, preventative maintenance, and first line maintenance training courses to two members of each participating centre's maintenance team.
- The cost of any contracts available for the regular service and/or maintenance of the devices should be separately stated.
- Prices should include any discounts or special offers available, including those of any ex-demonstration units that may be available, but these should be separately detailed.
- The terms of guarantee must be stated.

5. Quality Control

Quality assurance and control of the Ultrasound component of INTERGROWTH-21st will be the remit of an Ultrasound Quality Unit (USQU) (see section 6 below).

In order to ensure ongoing quality control we will employ rigorous tools for training, assessment and certification under the supervision of qualified instructors. Intra-observer and inter-observer measurement errors will be assessed during the training course before initiating the study.

Ongoing quality control and adherence to protocols will be performed. This will include quality assessment of images, random evaluation and repetition of ultrasound

measurements, and assessment of collected data. It is envisaged that a limited number of experienced obstetric ultrasonographers will form the research teams at each study site. They will be provided with standardized equipment and the ultrasound protocol describing all measurement techniques, protocols and procedures for training and clinical use.

5.1 Initial training and standardization of all study 2D-ultrasonographers against an international lead ultrasonographer, and assessment of local conditions

5.1.1 Initial training, assessment and standardization of ultrasonographers

It is recognized that the local ultrasonographers will already have a high standard of training. Therefore, the goals of initial training are:
- To ensure all ultrasonographers are familiar with equipment to be used in the study, including the ultrasound machine, the SOPs, the automated recording of ultrasound measurement values, software and entry of data.
- Ensuring standardization by training, assessment and certification. The schedule for this is as follows:
  1. Individual and group theoretical training (Coordinating Unit)
  2. Hands-on ultrasound training (Coordinating Unit)
  3. Submission of a log of 10 measurements of each parameter. These measurements will be repeated three times from three different images (to assess intra-observer variability), and repeated offline by the external sonographer (to assess inter-observer variability):
  4. Measurements should, on average, score >75% of the max score (i.e. 4 and 3 respectively, see below (5.2.1)
  5. No more that one out of the ten repeated measurement by the same operator should vary of more than two standard deviations (SD) of the measurement error of each given parameter at the given GA.
  6. No more than two out of the ten repeated measurements by a different operator should vary by more than two SD (proxy for random error) and there should not be a systematic bias of more than 0.5 SD.
  7. If these criteria are not fulfilled, certification will not be given.
  8. If the criteria are fulfilled certification of competence will be given prior to the start of the study

5.1.2 Pilot test-retest study

Similar to the anthropometry component of the study, quality control of ultrasound measurements will be primarily based on the comparison of repeat measurements by the same or different observers. Routine data collection will be guided by a system of maximum allowable differences between replicates. However, in contrast to anthropometry, for ultrasound measurements not much is known about what the maximum allowable differences should be for the different fetal measures at different stages of pregnancy. It is not known how exactly fetal size and GA influence measurement error in ultrasound. In addition, it is not clear how much measurement error in ultrasound is influenced by expertise and how expert-level and non-expert-level measurement errors compare. A pilot test-test study is therefore proposed to determine the Technical Errors of Measurement at both levels of expertise and trends to bias (in comparison with expert), specifically for:

- CRL between 9^{+0}-14^{+0} weeks
- BPD, HC, FL, AC in the different GA intervals that will be used in the study:14-18, 19-23, 24-28, 29-33, 34-38 and 39-42 weeks

This pilot study will be carried out in the unit at Oxford. It will involve one or several other obstetricians who regularly do ultrasound measurements in the same hospital. All observers

should be familiar with the measurement protocol of the study and should have received at least one training session by the LU prior to the start of the pilot study. The same equipment must be used as the one that will be used in the main study. For each of the seven GA intervals (from 9^{+0}-14^{+0} to 39-42 weeks) a number of 10 or 15 women will be enrolled over a 2-month period (November – December 2008).

It is proposed that each woman undergoes 4 independent assessments in a random order, two assessments by the LU and two by another ultrasonographer in the hospital. Each assessment will consist of an ultrasound session during which all the measurements needed for the GA interval will be performed in a random sequence. For each measurement a number of metadata (factors potentially modifying measurement error) could be collected so that the pilot study of itself becomes more publishable. All observers should ideally be blinded to the measurement values they obtain by automatic transfer of values to the online system.

This pilot study will yield GA-specific Technical Errors of Measurement and other statistics of inter- and intra-observer reliability and bias. These, in turn, will be used to define the maximum allowable differences for each fetal measure at each age that will be used to: (1) evaluate success of initial training (2) identify the need for re-measurement in the routine data collection, and (3) interpret data quality statistics from the standardization exercises and the random re-measurements.

5.1.3 Assessing local factors: the initial site visit

Prior to the start of the study, an initial site visit will be performed by a member of the USQU. The aims are to ensure:

- Preparation of the local data quality control activities, which will involve
  - Identification of a local supervisor of ultrasound data quality control activities. This will often be the ultrasonographer with the best data quality statistics and closest in performance to the LU during the initial LU visit. The person will have good organizational and team working skills and be competent in using the required software. The local supervisor should be instructed and trained to do the following tasks:
    - Conduct refresher sessions every two months, and document observer performances
    - Conduct data quality control re-measurements on a random sample of routine measurements
    - Observe as many routine measurements as possible performed by all site ultrasonographers
    - Extract and analyze, in collaboration with a LU, data from routine standardization sessions and from random re-measurements done at the site.
    - Liaise with LU about problems with ultrasonographers, SOPs and data quality statistics
  - Identification of a local data management coordinator who will liaise with the overall study coordinator.

5.2 Quality control of measurements in the FGLS: Monitoring and feedback of performances of ultrasonographers and identification of needs for retraining

5.2.1 Qualitative quality control
All measurements taken by ultrasonographers will be qualitatively controlled and scored by the Research Fellow.
Scores will be given according to the described scheme by *Salomon et al[11]*. Each specific criterion scores one point; thus, the maximum score is 6 points for BPD/HC; 6 points for AC; and 4 points for FL.
Any ultrasonographer with more that 10% of images rejected in a given period will have his/her certification withdrawn and will undergo re-training.

Objective scoring system for still images (modified from Salomon et al[11])

BPD/ OFD/ HC	AC	FL
Symmetrical plane	Circular plane	Both ends of the bone clearly visible
Plane showing thalami	Image shows the stomach bubble	<45∘ angle to the horizontal
Cavum septum pellucidum 1/3 along midline echo	Image shows umbilical vein along 1/3 of the abdomen	Femoral plane occupying at least 30% of the total image size
Cerebellum not visible	Kidneys not visible	Calipers placed correctly
Fetal head occupies at least 30% of the total image size	Abdomen occupies at least 30% of the total image size	
Calipers and dotted ellipse placed correctly	Calipers and dotted ellipse placed correctly	

In case of clearly incorrect caliper placement, images will be rejected and measurement excluded. If the image scores half or less of the maximum score (i.e. 3 points or 2 points, respectively), measurements will also be excluded.

5.2.2   Quantitative quality control: Intra-observer reliability

Intra-observer reliability will be prospectively assessed based on the three concealed measurements taken routinely. No more that **one out of the ten** repeated measurement by the same operator should vary by more than two SD of the given parameter at the given GA.

If this is not fulfilled, certification is cancelled and the sonographer identified as requiring re-training. Feedback and discussion between USQU and site supervisors will follow in order to highlight the need for re-training and to carry out re-training.

5.2.3 Quantitative quality control: Random re-measurement on 2D images

In order to assess correct caliper placement, a random sample of 10% of all measurements will be re-measured on still images by the Clinical Research Fellow / LU. This will allow production and evaluation of inter-observer reliability and bias (against the LU) for each observer based on the random re-measurements. No more than **one out of the ten** repeated measurement by a different operator should vary by more than two SD and there should not be a systematic bias of more than 0.5 SD.

If this is not fulfilled, certification is cancelled and the sonographer identified as requiring re-training. Feedback and discussion between USQU and site supervisors will follow in order to highlight the need for re-training and to carry out re-training.

5.2.4   Quantitative quality control: Random use of 3D images to re-measure planes

In order to assess correct image acquisition, a random sample of 5% of all 3D volumes will be used to acquire the ideal measurement plane by the Clinical Research Fellow / LU. This will allow evaluation of inter-observer reliability and bias (against the LU) for each observer based on the random re-measurements. No more than **two out of the ten** repeated measurement by a different operator should vary by more than two SD (proxy for random error) and there should not be a systematic bias of more than 0.5 SD.

If this is not fulfilled, certification is cancelled and the sonographer identified as requiring re-training. Feedback and discussion between USQU and site supervisors will follow in order to highlight the need for re-training and to carry out re-training.

5.2.5 Site standardization exercises.

Local test-retest exercises involving all site observers are an opportunity for refreshing the understanding of the protocol, better focus on the protocol, re-training and better comparison of observers as each observer measures the same subjects, which is not the case for routine measurements and for the random quality control re-measurements.

- o Organize periodic production and evaluation of data quality statistics for each observer based on the measurements taken during the standardization sessions
- o Feedback and discussion between USQU and site supervisors about data quality statistics from the standardization sessions; identification of need for re-training; organize and carry out re-training
- o These exercises will be done on 10 patients every 6 months, or more frequently depending on UQDU recommendations

5.3 Optimizing the data systems to accommodate the needs for quality control

The data entry system should allow incorporating measurement values and data from quality control and standardization exercises and be able to identify these as such.

- o The data system could select a random sample of non-supervisor routine measurements for QC re-measurement. Ultrasonographers will be unaware which measurements will be re-measured when they do their measurements.
- o Data extraction routine for monitoring of data quality statistics of entire study, sites, and individual ultrasonographers
- o Establishment of a research dataset on determinants of ultrasound data quality, to be extracted from the main database.
  - ▪ Possible factors causing variability in ultrasound measurements with a given type of instrument may include, among others:
    - • Parity
    - • BMI
    - • GA, size
    - • Presentation / position of fetus
    - • Sex
    - • Liquor volume
    - • Order of measurements
    - • Observer factors
    - • Measurement setting factors

5.4 Reporting, analysis and publication on ultrasound data quality
*5.4.1 Central reporting*

© University of Oxford

The USQU will produce 3-monthly data quality statistics based on standardization sessions, QC re-measurements and routine replicates. This may include, as appropriate the production of statistics and trend plots of:

- Intra-observer reliability of individual ultrasonographers
- Inter-observer reliability of individual ultrasonographers against LU
- Site bias against LU
- Site intra-observer reliability
- Site inter-observer reliability against supervisor
- Comparison among sites
- Proportion of failed checks on maximum allowable differences and on range checks

B. Publication plan

- Publications on ultrasonographic training and data quality will be integrated into the study's overall publication policy
- Possible topics include:
  - Paper on ultrasound data quality in INTERGROWTH 21st
  - Research questions using the research dataset on ultrasound data quality
    - Determinants of observer bias in ultrasonography
    - Determinants of observer reliability in ultrasonography
    - Differences between concealed and revealed measurements
    - Difference between 2D and 3D measurements
    - Differences in data quality statistics obtained from training sessions, routine data collection, quality control, and standardization sessions

Figure: Summary of quality control
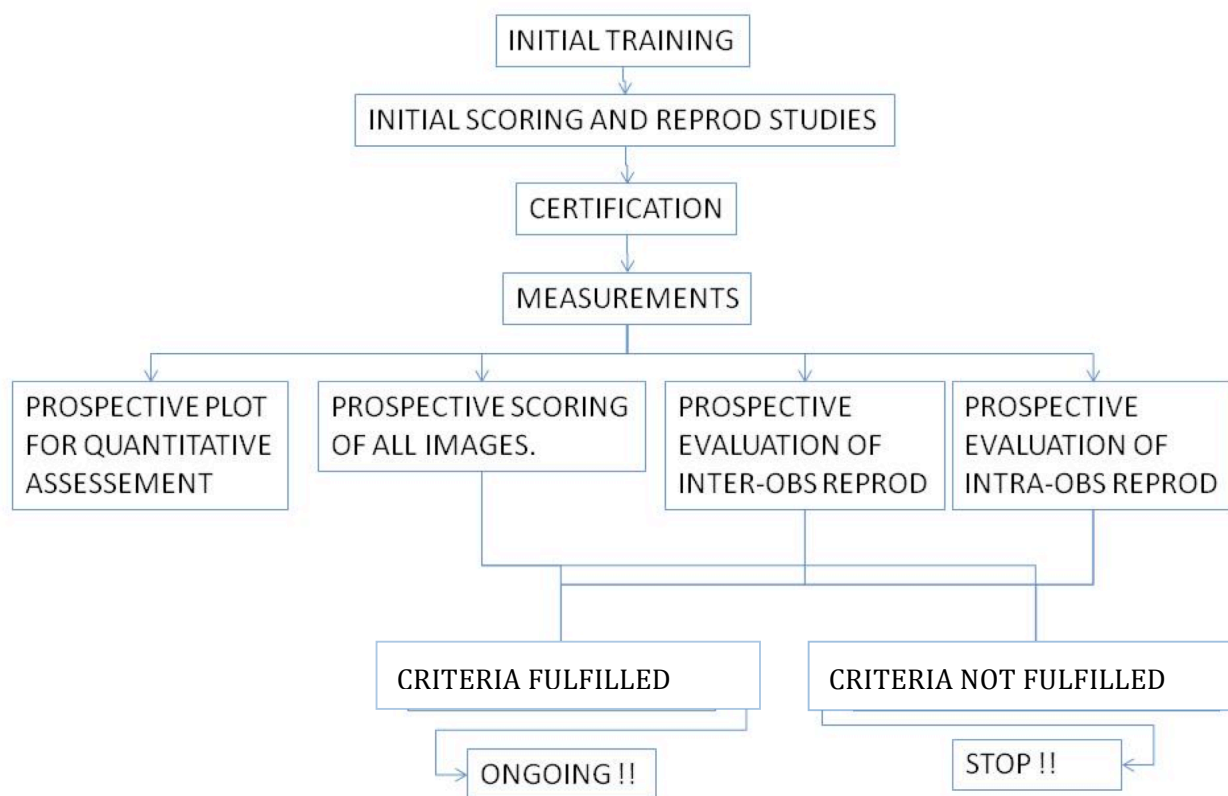
6. Ultrasound Data Quality Unit

Quality assurance and control of the Ultrasound component of Intergrowth-21st will be the remit of a Ultrasound Quality Unit (USQU). USQU will be coordinated by Aris Papageorghiou who will be assisted by External Ultrasound Experts (Laurent J Salomon and Shaida Zaidi) a Clinical Research fellow (Study protocol p.34). This unit will co-ordinate with the anthropometric quality control unit headed by Dr Jan van den Broeck, and both will follow similar protocols.

6.1 Proposed activities of the USQU

The USQU will be responsible for:

1. Development of standard operating procedures (SOP)
2. Initial training, assessment and certification of ultrasonographers
3. Standardizing all study 2D-ultrasonographers against a lead ultrasonographer (LU)
4. Monitoring and feedback of on-site standardization levels and performances of individual ultrasonographers; identification of needs for retraining
   a. Quality control of routine measurements
   b. Random sample of replicate measurements
   c. Site standardization exercises
5. Optimizing the data system to accommodate the needs for quality control
6. Reporting, analysis and publication on ultrasound data quality
   a. Central reporting
   b. Establishment of an analysis dataset on ultrasound data quality, to be extracted from the main database
   c. Publication plan

© University of Oxford

6.2 Ultrasound protocol and standard operating procedures (SOP).
This SOP includes procedural guidelines on:
- Patient selection
- Preparation of subject for measurement
- Sequence of measures and replicates; blinding to previous measurement values
- Measurement technique
    o Positioning of subject
    o Handling of caliper, identification of planes and landmarks, marking distances, recording and transfer of results
    o Entry of measurement metadata
- How to deal with special fetal positions and other unusual circumstances; circumstances under which one should delay or forego the assessment or part of the assessment;
- Release of measurement values and GA calculations for clinical purposes
- Preparation of measurement setting and instruments; linkage with computer and database
- Ultrasound device and accessories: maintenance, transport, storage, calibration checks

6.3 Unanticipated changes to this protocol / SOP
Despite all efforts, it is likely to be impossible to anticipate all possible eventualities in a study of this size. Therefore, changes to this protocol will be possible. These will be proposed to the Steering Committee by the USQU, and incorporated if approved.

References

1. The International Fetal and Newborn Growth Consortium for the 21st Century INTERGROWTH-21st Study Protocol. Oxford University, 2008

2. Cetin I, Cozzi V, Antonazzo P. Fetal development after assisted reproduction - a review. Placenta. 2003 Oct;24 Suppl B:S104-13

3. Jacob S, Moley KH. Gametes and embryo epigenetic reprogramming affect developmental outcome: implication for assisted reproductive technologies. Pediatr Res. 2005 Sep;58(3):437-46

4. Robinson HP. Sonar measurement of fetal crown-rump length as means of assessing maturity in first trimester of pregnancy. Br Med J. 1973 Oct 6;4(5883):28-31

5. Robinson HP, Fleming JE. A critical evaluation of sonar "crown-rump length" measurements. Br J Obstet Gynaecol 1975;82:702-10.

6. British Medical Ultrasound Society. Fetal size and dating: Charts recommended for clinical obstetric practice. February 2007 (Revised: January 2008)

7. Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 2. Head measurements. Br J Obstet Gynaecol 1994; 101: 35–43.

8. Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 3. Abdominal measurements. Br J Obstet Gynaecol 1994; 101: 125–131.

9. Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 4. Femur length. Br J Obstet Gynaecol 1994; 101:132–135.

10. Snijders RJM, Nicolaides KH. Fetal biometry at 14–40 weeks' gestation. Ultrasound Obstet Gynecol 1994; 4: 34–48.

11. Salomon LJ,. Bernard JP,. Duyme M, Doris B, Mas N, Ville Y. Feasibility and reproducibility of an image-scoring method for quality control of fetal biometry in the second trimester. Ultrasound Obstet Gynecol 2006; 27: 34–40

12. International Society of Ultrasound in Obstetrics & Gynecology Education Committee. Sonographic examination of the fetal central nervous system: guidelines for performing the 'basic examination' and the 'fetal neurosonogram'. Ultrasound Obstet Gynecol. 2007 Jan;29(1):109-16. Links

13. Royal College of Radiologists: Standards for ultrasound Equipment (2005)

Appendix D: NEWBORN, PRETERM AND MATERNAL Anthropometric Techniques, Equipment and Standardization and Data Quality Assurance and Control Plan

1. Background

The three components of INTERGROWTH-21st are a fetal longitudinal growth study (FLGS), a newborn cross-sectional study (NCSS) and a preterm postnatal follow-up study (PPFS). All components focus essentially on aspects of growth and will perform anthropometric measurements that will be based on ultrasound before birth and on classical anthropometry after birth. As in all clinical-epidemiological research, data quality in INTERGROWTH-21st depends on: (1) Accurate and reproducible measurements of key variables, and (2) Preservation of data integrity after measurement, from recording until analysis. Key variables in anthropometry are age, sex and variables for body measures. Post-measurement data integrity depends on data systems and management.

This Appendix describes the overall approach to data quality assurance and control in INTERGROWTH-21st. Key functions of quality assurance/control of measurements include training, standardization and calibration of measurement procedures (including recording), monitoring of measurers' performance and data cleaning. Key to maintaining data integrity in the post-recording stage is a secure database as well as standardizing and documenting all steps in the data flow.

Most of these quality assurance/ control functions will be the responsibility of two Data Quality Units that will be established within INTERGROWTH-21st. Quality assurance and control of the Ultrasound component of Intergrowth-21st will be the remit of a Ultrasound Data Quality Unit (UDQU). UDQU will be coordinated by Aris Papageorghiou who will be assisted by External Ultrasound Experts (Laurent Salomon and Shaida Zaidi) and a Clinical Research Fellow. Quality assurance and control of the postnatal components of INTERGROWTH-21st will be the responsibility of an Anthropometric Data Quality Unit (ADQU) and will be coordinated by Jan Van den Broeck and three or four regional lead anthropometrists (RLA).

2. Selection and initial training and standardization

Each INTERGROWTH-21st study site will work with two to four ultrasonographers and a similar number of postnatal anthropometrists. Although some will already have a high standard of training, all will receive training to follow procedures according to the detailed measurement and standardization protocols of the study and up to a level that approaches or surpasses that of designated expert trainers.

2.a. Selection of observers

All observers will be clinical hospital staff with prior experience in ultrasound or infant anthropometry. In addition they will possess good fine motor skills and vision and a methodological and disciplined approach to task completion. They must speak the local language and preferably English as well.

2.b. Measurement and standardization protocols

The most important question-based measurements, relevant to all study components, are the LMP and the regularity of menstrual periods. This is because these questions provide essential information to calculate the key variable - gestational age. It is vitally important that these questions are standardized and answered as accurately as possible. In classical post-natal growth studies the accuracy and precision of the results tend to be free or nearly free of the influence of measurement error in age determination. This is different in pre-natal growth studies where the results are under the influence of errors in both body measurements and gestational age determination. The influence of the latter will be reduced to a minimum by restricting participation in the FGLS to women with regular menstrual

© University of Oxford

periods and a certain LMP, which is consistent with the CRL. For the same reason, CRL is among the essential body measurements in INTERGROWTH-21st requiring the highest level of training and standardization.

Each body measurement value obtained, whether in research or in clinical practice, is the end result of a complex interaction between the measurement setting, the subject (usually with accompanying relatives), the instrument(s), and one or more observers who measure and record values. Importantly, numerous factors other than the body dimension itself have the potential to influence a measurement result. These factors can lead to biased and imprecise measurement values and thus their influence needs to be avoided or minimized. This is the main purpose of standardization protocols. Specifically, these protocols aim at reducing the influence of inaccuracy and imprecision of instruments, sub-optimal measurement settings, undue body movements and postures, hydration, deficiencies in the observer's technical skills (experience, protocol knowledge), fitness and mood and errors in data recording.

In INTERGROWTH-21st, separate measurement and standardization protocols will be developed for ultrasound and postnatal anthropometry but all protocols will include procedural guidelines on:
- Choice of equipment; maintenance, transport, storage and calibration procedures. All study sites will use the same type of certified, accurate and highly precise instruments that fulfil specified technical minimum requirements.
- Subject recruitment (inclusion-exclusion criteria and their identification), enrolment and preparation.
- Preparation of measurement setting and instruments for measurement sessions. Essential to all body measurements is good visibility of the scales or images shown by the measuring equipment. Equally important are guidelines to ensure comfort of the subjects and the observers by eliminating or minimizing any influences that may affect concentration on, and implementation of, tasks.
- Schedule and sequence of measurements; ideally the entry system will allow sequence of measurements to be determined randomly for each session; depending on the case, a measurement session may involve compulsory measurements only or additional body measurements, re-measurement after failed range check yes/no, quality control re-measurement by local supervisor yes/no, additional collection of measurement metadata yes/no; all measurements will be made in duplicate by the same observer; maximum allowable differences between the two measurements are checked and may lead to flagging for re-measurement. If the values fall within the allowable differences they will be averaged and this will be taken as the final value; this method increases accuracy
- Blinding to previous measurement values of the same subject is important to ensure that measurement values are not influenced by subjective observer expectations
- Measurement technique in the strict sense
  - Subject positioning; positioning of observers
  - Roles of assistants, mothers and relatives
  - Handling of instrument; identification of planes and landmarks
  - Markings, readings, recordings
- Transfer of measurement results into database
- How to deal with special positions and other unusual circumstances; circumstances under which one should delay or forego the assessment or part of the assessment;
- Proximity of entry software. With ultrasound, as opposed to the other measurements, data entry software and measurement instrument will be directly linked. It is important to note that for measurements that require manual entry of visual or auditory information (postnatal anthropometry and questions) computer and entry software must also be accessible in direct proximity of the measured subject

otherwise the delay between perceiving the information and recording it may become too long, and may undo the advantages of direct entry.

2.c. Initial training and standardization

Initial training will be organized for all observers to ensure that they are familiar with instruments and questionnaires used in the study, with the measurement and standardization protocols and with study software and automated recording of data. Standardization will be achieved through training, assessment and certification of competence. To this end experts from coordinating centre and data quality units will ensure that all observers will be certified prior to the start of the study. This will involve site visits by experts for:

- o Individual and group theoretical training
- o Hands-on training on the measurement and standardization protocols, including those regarding questionnaire delivery, gestational age determination and body measurements
- o Initial test-retest exercises for ultrasound and for postnatal anthropometry. In each test-retest exercise observers will do independent duplicate measurements of 15 subjects. The same subjects will also be measured by the visiting expert who will serve as a 'gold standard'. The results will be used for estimation of baseline observer and site reliability and bias. These estimates will form a basis for data quality monitoring during the data collection phase of the study
  - Bias of individual observers against expert
  - Intra-observer reliability of individual observers
  - Inter-observer reliability of individual observers against expert
  - Site bias against expert
  - Site intra-observer reliability
  - Site inter-observer reliability against expert
- o Local adaptations to standardization protocols

During the initial site visits two local supervisors/instructors of data quality control activities will be identified, one for ultrasound and one for postnatal anthropometry. These will often be the observers with the best data quality statistics and closest in performance to a visiting expert. These persons will have good organizational and team-working skills and be competent in using required software. During the expert visits the local supervisors will be instructed and trained to do the following quality control tasks in their respective domains (ultrasound or postnatal anthropometry):

- o Conduct refresher training sessions and test-retest studies every three months using similar procedure as above, except that only 10 subjects will be measured; once or twice during the study an external expert will revisit the site to participate with these follow-up test-retest exercises
- o Use the measurement values of the test-retest studies to document observer performances including individual ad site biases against the supervisor and inter-observer reliabilities against the supervisor
- o Conduct data quality control re-measurements on a 10% random sample of routine measurements
- o Observe as many routine measurements as possible of all site observers
- o Liaise with their respective data quality unit about problems with availability of observers, with measurement and standardization protocols and with data quality statistics

3. Quality control during data collection

There will be several types of quality control activities during the data collection phase of the study.

3.a. Quality control of routine measurements:
These will include the following activities:
- Frequent observation of measurements by local site supervisor
- Checks of distributions of each measurement value: cross-sectional based on attained size and longitudinal based on increments; existing growth references will be used to define the normal ranges; these range checks will be automated and immediate and may lead to flagging for re-measurement; missing data will also be flagged
- Immediate automated checks of maximum allowable differences between repeat measurements; maximum allowable differences will be determined on the basis of what is considered normal intra-observer reliability of experienced observers
- Feedback and discussion between quality control units and site supervisors identification of need for re-training; organize re-training
- Random quality control re-measurements. A random sample of 10% of all measurements will be re-measured by the local site supervisor; In addition, for ultrasound a random 10% of stored images will be subjected to rescoring by an expert from the UDQU; Production and evaluation of inter-observer reliability and bias statistics for each observer based on the random re-measurements
- Terminal digit preference of observers will be monitored for measurements of continuous variables that require manual entry of result values
- Continuous feedback and discussion between data quality units and site supervisors; identification of need for re-training; carry out re-training
- Data cleaning (including outlier and missing data detection) and study progress monitoring by the data management team

3.b. Site standardization exercises
These test-retest exercises involving all site observers are an opportunity for refreshing the understanding of the protocol, better focus on the protocol, discussion, re-training and better comparison of observers as each observer measures the same subjects, which is not the case for routine measurements and for the random quality control re-measurements
- These standardization exercises will be organized by the local supervisors every three months and may or may not be attended by a visiting expert from the data quality units; an expert will participate at least once in the data collection phase
- In each standardization session all observers of the site will do independent duplicate measurements on a total of 10 subjects. The same subjects will also be measured by the local supervisor and sometimes by visiting expert who will serve as 'gold standards'.
- The data quality units, in collaboration with the local supervisors, will organize periodic production and evaluation of data quality statistics for each observer and each site based on the measurements taken during the standardization sessions
  - Bias of individual observers against local supervisor and/or visiting expert
  - Intra-observer reliability of individual observers
  - Inter-observer reliability of individual observers against local supervisor and/or expert
  - Site bias against local supervisor and/or visiting expert
  - Site intra-observer reliability
  - If applicable: inter-observer reliability and bias between local supervisor and visiting expert

- o There will be feedback and discussion between the data quality units and the site supervisors about data quality statistics from the standardization sessions; identification of need for re-training or training of new observers; organization and carrying out of (re-)training

4. Reporting of data quality

The data quality unit will produce 3-monthly reports describing data quality statistics for sites and observers, based on standardization sessions, random quality control re-measurements, routine replicates and, as applicable to specific study components, ultrasound quality scores, QUSUM stats and terminal digit preferences. This will include the production of trend plots of the same statistics.

Publication of scientific papers on the data quality assurance procedures and actual data quality is part of the study's overall publication policy.

Appendix E: **Data Management: the Electronic System**

The data management process will be performed electronically, simultaneously with data collection, at each study site. Data entry will be implemented on a customized web-based system developed for the study. The system automatically carries out range and consistency checks for immediate correction and reviews percentages of missing/unknown values for each variable. This technology also allows a real-time audit of each site by a set of pre-defined reports included in the system (recruitment, data inconsistencies, etc) run periodically by the Project Data Manager. Additional modules for internal/external messenger, hosting of study documentation (protocol, guidelines, etc) and user administrator make up an integral data management solution for multicentre studies.

The INTERGROWTHth-21st website homepage will consist of an open part and a part protected by username and password. The open part will be used for publication of information, news etc. The protected part will contain the data entry application, a reporting module, an alert module, a descriptive statistics module, and an administration module. We are looking into the possibility of linking the application to the external ultrasound data sources.

Ultrasound measurements obtained during the follow-up visits will be electronically transferred from the ultrasound equipment to the web-based system to avoid transcription errors using a database format, agreed between ultrasound equipment manufacturer and system development group. Routine analyses will be carried out regularly for each variable to check digit preference and unusual frequencies of answers that may reflect poor understanding of the instructions. After the information is electronically transferred, the last measured will be made available to the attending staff for clinical use. The application includes an automated user and event logging system.

Data collected will be evaluated by the Data Management Unit for further quality control analyses and compliance with the study protocol, including the analysis of eligibility criteria, the timing of visits and drop-out rates. Measurements for the study taken from the equipment will have a facility to blind all fetal measurement values until they are submitted for the study. Following submission the last value will be printed for use in clinical care.

All other clinical and anthropometric data will be collected initially in a paper form entered onto the on-line system at each local institution by specially trained data management staff. Data should be entered into the system within one week of being collected facilitating the retrieval of possible missing data or other inconsistencies.

Appendix F: Policy and determination of responsibilities for INTERGROWTH-21st
1. Study Coordination and Management

The study's coordination and management, including data management and analyses, are the primary responsibility of the Oxford Maternal & Fetal Health Institute (OMFHI) within the Nuffield Department of Obstetrics & Gynaecology, University of Oxford, which will act as the study Coordinating Unit (CU). The Department of Nutrition of the World Health Organization (WHO) will collaborate with OMFHI in the study's overall implementation.

The study coordination and management will include the following activities:

1.1 Study Preparation

Coordination activities:
- Develop protocol with collaborators
- Conduct site visits for selection of study sites
- Selection of committee members and assign tasks to study committees
- Organize collaborators' meetings
- Establish communication procedures between sites and CU
- Administer funding base for coordination activities and study implementation
- Keep donor agencies informed about progress of study

Study materials:
- Prepare data collection forms and supporting documents
- Prepare operation manuals
- Prepare measurement and standardization protocols including Quality Assurance
- Select and purchase equipment
- Prepare data collection forms and other study materials as required

Data processing and management systems:
- Conduct feasibility study
- Negotiate and implement an electronic data management system
- Set-up system for on-line data entry and validation
- Define electronic monitoring reports to be produced
- Pilot data collection systems and monitoring reports

Statistical analysis:
- Develop statistical analysis plan

1.2 Study Implementation

Coordination activities:
- Assist study sites in local adaptation of study materials
- Submit protocol to Ethics Committees
- Distribute data collection forms and other study materials to centres
- Develop standard set(s) of slides for presentations about the study
- Establish training procedures for research staff

1.3 Study Conduct

Coordination activities:
- Coordinate study's day-to-day activities
- Monitor study progress through site visits and on-going review of data entry
- Communicate with local investigators; supporting units; Steering Committee (SC) and Advisory Committee; and other collaborators (e.g. UN agencies, donors)

- Prepare electronic study newsletter
- Convene regular meetings of staff at CU
- Organize coordination meetings of local Principal Investigators; Steering Committee; site visits; exchange of relevant staff between sites, etc.

Data management and statistical analysis:
- Monitor data collection procedures at study sites and regularly advise SC about all data received and processed
- Validate data and produce queries about inconsistent and/or apparently incorrect data
- Send data queries to study sites so that errors are corrected at collection phase
- Update master file using corrections from validation checks and/or answers to queries
- Evaluate electronic monitoring reports by study site based on recruitment rates, drop-outs, data completeness, quality assurance checks, etc.
- Maintain communication channels between study sites and supporting units relating to queries
- Inform SC whenever a study site requires retraining in data collection or is failing to respond to queries and implement corresponding actions
- Conduct interim statistical analyses of data in accordance with analysis plan approved by SC
- Conduct final statistical analyses of data in accordance with plan approved by SC
- In collaboration with SC, coordinate preparation of articles for publication in peer-reviewed, scientific journals etc.

2. Study Committees

The following Committees will oversee the implementation of the study:

2.1 Steering Committee (SC)

The SC will consist of the following:

- Professional staff from CU and its supporting units
- Principal Investigators (one per site)
- Selected senior scientists
- Representatives from Department of Nutrition and Department of Environmental Health, WHO, and other related Institutions as considered appropriate

The SC will meet regularly by conference call or face-to-face meetings to review the study's progress and discuss substantive issues that arise from the study's implementation and conduct. Its role is to make managerial decisions that affect the conduct of the study; to ensure the comparability of data from the different sites, and to resolve problems that arise in running this complex project.

The SC will have an **Executive Committee (EC)** that will meet frequently to review study progress and problems, protocol divergences, and other substantive issues that may arise from the study's implementation. The EC will have overall administrative responsibility for the study's implementation. All changes or alterations to the protocol or issues related to the technical, financial or administrative conduct of the study must be approved by the EC. This Committee will also select the study sites; approve the continuing participation of the sites selected, and determine which data are to be included in the pooled, international data set. Every effort will be made to achieve consensus for all decisions but where this is not possible resolution will be by simple majority with a quorum of at least 5 members. The EC membership will be JV, SK, AN, DA, MdO, ZB, AP, the SC Chair and a Principal Investigator

selected by his/her peers. The Project Manager will be an ex-officio member.

2.2 Advisory Committee (AC)

- The AC will consist of senior scientists in the areas of Medicine, Epidemiology and Statistics, Obstetrics and Gynaecology, Paediatrics and Human Biology. The task of the AC will be to provide overall and strategic advice to the SC.

3. Data Management

Electronic data management will be contracted to Medscinet, a private company that has provided similar services in previous multicentre studies we have conducted. It has extensive experience of conducting large multicentre trials and observational studies, and it is recognized as a leading research organization in reproductive health. Although based in Sweden, the company has a UK base in the Department of Women's Health, St Thomas' Hospital, London. It will provide a full on-line data management system, as well as supervision and training support to the participating sites for the purposes of data entry, cleaning, file preparation, and study monitoring. All participating sites will be required to adhere to the study's data management system, which will be available on the study web page. We have successfully used this system in two previous multicentre studies involving sites included in INTERGROWTH-21st.

4. Ancillary Studies

Proposals for ancillary studies from one or more local Principal Investigators will be considered as long as a) they do not result in alterations to the main protocol and b) they do not impact adversely on subject recruitment and/or participation in the main study. All ancillary studies and/or additional data collection must be submitted to the EC for written approval before implementation.

A letter of intent must be submitted to the CU at the conceptual stage of the ancillary study. If approved by the EC, the local Principal Investigator(s) will be invited to submit a detailed proposal that should include the study's rationale, objective, methodology, expected outcomes and budget.

5. Site Monitoring

Study progress will be monitored by:

a) Monthly reports for each study site produced by the CU using data from the electronic data bank, which will be shared with the study sites and EC members.

b) Visits (the frequency as required by the progress of the study) to all study sites to ensure that the study is uniformly implemented (the timing to be decided by the CU in coordination with the EC and Anthropometric Quality Control Unit (ADQU). The activities to be performed during each visit include, but are not restricted to:

- Meeting site Principal Investigator and members of study staff
- Review of study procedures, use of data forms and related documents
- Observation of study personnel carrying out specific procedures
- Particular attention will be paid to standardization of data collection (e.g., adherence to study protocol, and measurement/standardization protocols)
- Review of coverage and participation rates, including characteristics of participants and non-participants, and reasons for loss to follow-up

- Observation of data validation procedures
- Physical walk-through of certain procedures (e.g., screening to determine subject eligibility or taking of informed consent)
- Conversations with key support personnel to assess their practice with regard to data collection
- Inspection of study facilities and subject study files

c) Exchange visits by field supervisors, regional and global meetings of field directors to facilitate standardization of procedures across sites.

6. Dissemination of Results, Data ownership, Publications and Authorship

6.1 Dissemination of Results

The results of the study will be disseminated as widely as possible, in publications and the mass media. Activities will include:

- Publication in peer-reviewed journals of all papers reporting primary and secondary outcomes based on the pooled international data set
- Publication of papers and editorials in regional journals that are published in local languages, such as Bulletin of PAHO (Spanish), Chinese, French or African Journals
- Electronic study newsletter
- Publication in periodicals from other collaborating institutions.
- Presentations at local, regional and international meetings
- Publication of a book including all collected information (e.g., growth curves) and detailed methodological issues unlikely to be included in reports to peer-reviewed journals. The Executive Committee in coordination with Local investigators will take responsibility for this task.

6.2 Data Ownership

All data derived from the INTERGROWTH-21st multicentre study will be the property of the Oxford Maternal and Fetal Health Institute (OMFHI) of the Nuffield Department of Obstetrics and Gynaecology at the University of Oxford. Individual investigators will share the ownership of their site-specific data sets with the OMFHI. After publication of the growth curves and the main papers related to the secondary objectives, the pooled data set will enter the public domain at a time to be determined by the SC. The anonymity of all study participants will be ensured.

6.3 Publications and Authorship

All papers reporting the new growth charts (Primary Objective, components 1-3), based on the pooled international data set (the growth curves) will be published under corporate authorship (International Fetal and Newborn Growth Consortium) The address for correspondence will be the Oxford Maternal & Fetal Health Institute (OMFHI), Nuffield Department of Obstetrics & Gynaecology, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

For these publications, the names of the members of all committees and study units will be listed at the end of the paper (see Lancet 2006;367:1819 for an example of the format) plus their individual contributions to the study (e.g., protocol development, study coordination, data management, data analysis, manuscript preparation, etc) as required by the individual journal.

Sites participating in the multicentre study will be listed alphabetically by country; the local Principal Investigator will be responsible for selecting the authors within each country. It is understood that authorship, within each site, will be offered to those who have made a substantial contribution to the study. Within each site, the order of the author's names will be the responsibility of the local Principal Investigator. The SC will provide a list of people to be mentioned in the standard Acknowledgments, once it has been agreed what type of support will be acknowledged. All papers arising from the study should include such a list of acknowledgements. The University of Oxford and WHO Nutrition will give certificates of collaboration to doctors, nurses and local staff who contributed to the study but whose names cannot appear in publications.

For all other papers arising from the study (except the main growth charts described above), including those relating to secondary objectives, a modified system of corporative authorship will be used, in which SC members will be listed by name according to an agreed rotation based on personal interests and leadership in the preparation of the paper. These names would be followed by "for the International Fetal and Newborn Growth Consortium" or similar statement as agreed by the SC. The authorship order will rotate based on a system agreed by the SC, i.e. initial lottery and rotation thereafter or some other system.

The study will have a **Publications Review Committee (PRC)** consisting of the EC members, a Local Principal Investigator and one or two *ad hoc* Senior Scientists invited according to specific needs. The PRC will be responsible for:
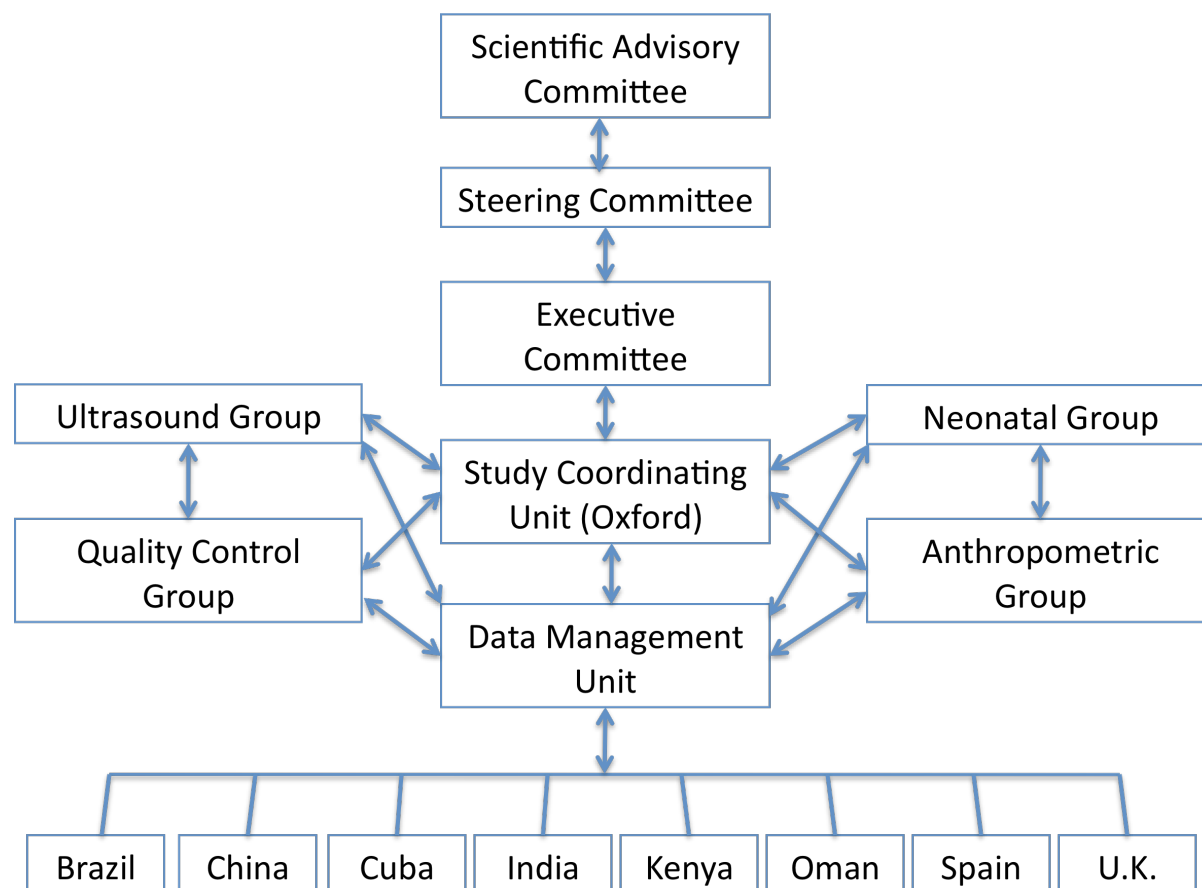
- Preparation of a list of expected publications on the primary and secondary outcomes of the study, with the authorship strategy for each set according to the rules above which are to be approved by the SC.
- Setting-up writing committees for the primary outcome, corporate authorship papers and a review mechanism by the SC of the final drafts that are submitted for publication.
- Reviewing and approving all papers sent for publication with the goal of maintaining internal consistency of material and methods, as well as authorship policies.
- Reviewing and approving all site-specific analyses proposed by local Principal Investigators.
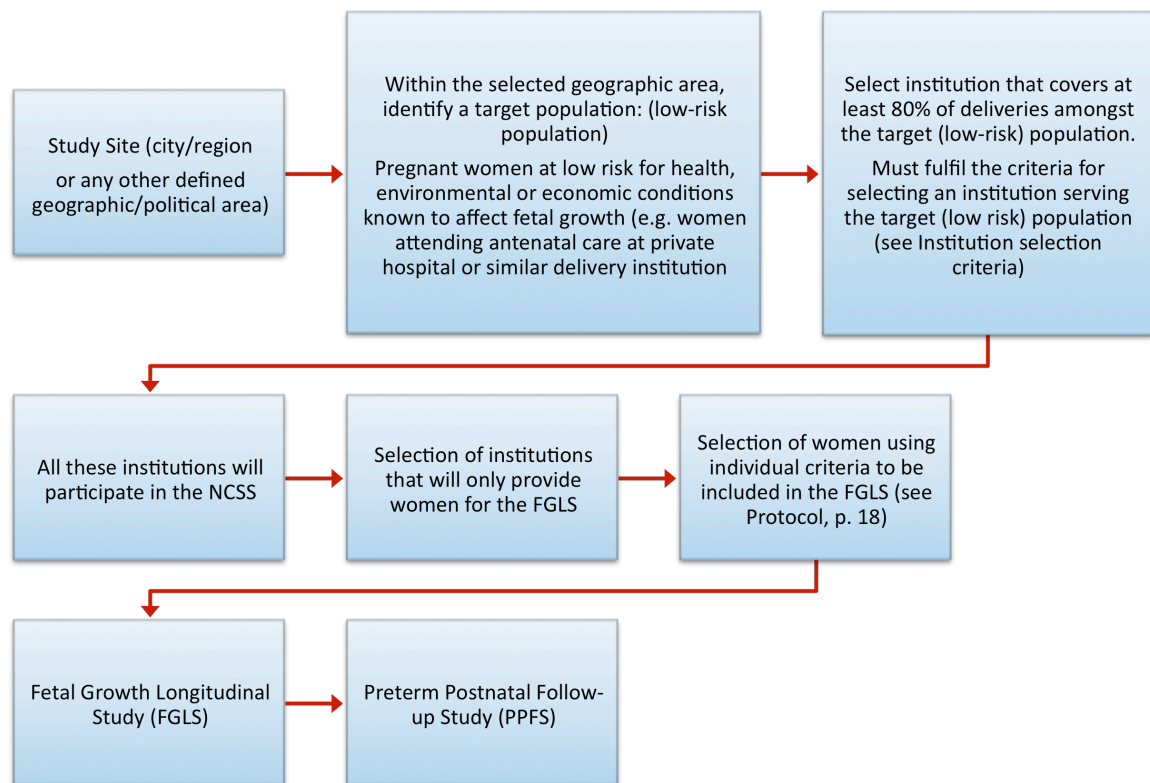
The following more general rules will apply:

- Individual sites will not be allowed to publish analyses of country-specific data or subsets of sites that, in the EC's assessment, have the potential to be misconstrued as standard data for particular nations or population groups.

- However, individual sites and groups of sites are encouraged to publish other analyses of data based on their site-specific data. These analyses must be cleared by the EC before preparation. The order of author's names in publications based on site-specific analyses will be the responsibility of the local Principal Investigator. The EC will provide the Acknowledgments and funding sources, in a standard format.

- The CU is encouraged to publish methodological and conceptual papers that describe the methods and procedures used in the study. The order of authors' names for these publications will follow the modified corporative authorship model including the SC as well as the CU members that lead the preparation of the papers. The CU will provide the Acknowledgments in a standard format.

- Scientists are encouraged to present the study at scientific conferences and meetings. When scientists are invited to international or regional meetings to present

the study, the CU should be informed and prior written approval is required with details of the type, venue and organizers of the meeting. The CU will keep an archive of all materials presented at meetings and make them available at the study's web page.

- Press enquiries will be honoured unless there are some operational or scientific reasons for withholding information. Requests for interim results or other details arising during the study which if honoured are likely to have an adverse effect on the study, will be denied. Only one individual will be authorized to interact with the press (the local Principal Investigator or public relations officer of his/her institution) at every site in coordination with the CU. In the multicentre study context, the EC will respond to queries concerning the overall study design or results. The SC will define the type of queries that may be answered locally and those that must be referred for response to the EC. Publicity concerning study results in preparation will be avoided.

Appendix G: Management Structure

Appendix H: Selection of Study Sites Flow

Study Site (city/region or any other defined geographic/political area)

→

Within the selected geographic area, identify a target population: (low-risk population)

Pregnant women at low risk for health, environmental or economic conditions known to affect fetal growth (e.g. women attending antenatal care at private hospital or similar delivery institution

→

Select institution that covers at least 80% of deliveries amongst the target (low-risk) population.

Must fulfil the criteria for selecting an institution serving the target (low risk) population (see Institution selection criteria)

All these institutions will participate in the NCSS

→

Selection of institutions that will only provide women for the FGLS

→

Selection of women using individual criteria to be included in the FGLS (see Protocol, p. 18)

Fetal Growth Longitudinal Study (FGLS)

→

Preterm Postnatal Follow-up Study (PPFS)

Institution Selection Criteria

a) LBW rate <10% and mean birth weight >3100g;

b) located at an altitude below 1600m;

c) perinatal mortality <20 per 1000 live birth;

d) mothers attending antenatal care in these institutions should plan to deliver in that or a similar hospital located in the same region;

e) >75% of mothers have attained an educational level greater than the locally defined cut-off point;

f) lack of known non-microbiological contamination such as pollution, radiation or any other toxic substances (this will be evaluated in collaboration with the WHO Department of Protection of the Human Environment);