Applications of path analysis and functional data analysis in a longitudinal, clinical cohort study of pregnant women and their neonates.

Kathrine Frey Frøslie

Thesis presented for the degree of Philosophiae Doctor

UiO : University of Oslo

Department of Biostatistics
Institute of Basic Medical Sciences
Faculty of Medicine
University of Oslo
Norway

Oslo, December 2013

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 1771

ISBN 978-82-8264-813-4

Cover: Inger Sandved Anfinsen.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Akademika Publishing.
The thesis is produced by Akademika Publishing merely in connection with the thesis defence. Kindly direct all inquiries regarding the thesis to the copyright holder or the unit which grants the doctorate.

Acknowledgements

This project was funded by the Norwegian Resource Centre for Women's Health, Oslo University Hospital. I am grateful to Professor Tom Tanbo, the former Centre Manager, for his faith in me and the project, and the present Manager, Siri Vangen, for encouragement and funding. I am also grateful to the Department of Biostatistics, University of Oslo, for providing an office and a stimulating and inspiring workplace.

I have had the most wonderful team of supervisors, and I am deeply indebted to all of you: to Marit B. Veierød for being so generous and inclusive, for your profound interest in every detail of my work, for always caring about my social life and family matters, for always being up-to-date, for sharing your knowledge and letting me build LEGO in your classes, for never letting go of your high standards, and for always making me perform at my best; to Jo Røislien, for thinking large and making me believe that some of these ideas were mine, for being liberatingly unconventional in your approach, yet very precise in your mathematical notation, for boosting my self-confidence, and for being utterly inspirational and fun to work with; to Tore Henriksen, for your good heart and kind spirit, for sharing your wisdom and your knowledge of human cells, physiology, metabolism, pregnancy and childbirth, of Darwin and life, and for your amazing ability to give me the impression that *I* contribute to *your* research, when it is actually the other way around; to Petter Laake, the mentor in this project, for your constantly relaxed and good mood, and for always offering calming advice whenever tricky issues needed to be solved. I am likewise indebted to Jens Bollerslev for your sharp academic view, your humour, and your enthusiastic and valuable contribution to my work – you have practically been my extra supervisor.

I am grateful to all my co-authors: to Camilla M. Friis for your energy, for presenting refreshing questions about inflammation and mediation, for challenging discussions and for sharing the first article with me; to Nanna Voldner for our good friendship, for letting me take part in your PhD project and for being such a patient colleague; to Kristin Godang, my favourite biochemist and constant provider of everyday amusement; to Thor Ueland and Elisabeth Qvigstad, for your encouragement and enthusiasm, and for sharing your knowledge about cellular mechanisms, physiology and clinical care.

The STORK study was designed by Tore Henriksen and Jens Bollerslev, and I truly admire and appreciate the large work in data collection performed by Nanna Voldner,

Camilla M. Friis, Marie Cecilie Paasche Roland, Kristin Godang and other STORK colleagues. I am also grateful to the women who participated in the STORK study.

During this period, I have had the pleasure to work in an environment full of interesting, intelligent, witty, good-natured, encouraging and, most of all, inspiring people. To all my colleagues and friends at the Department of Biostatistics, the Norwegian Resource Centre for Women's Health and Sunnaas Rehabilitation Hospital: thank you so much for distractions, good advice, joy and fun during hours and hours of coffee, lunch, dinner, scrabble, music rehearsals, walks and breaks. I'm the lucky one.

To my sweet nuisance parameters Ingeborg, Nils and Kjersti, utterly important for me and my progress, but not really the topic of my thesis: you are so patient! I really look forward to getting started with your "Tusen ting mamma skal gjøre når hun er ferdig med doktorgraden" -list, which you promised to write whenever I postponed something fun with "that will happen after my PhD". Now, let's fix the rabbits' lawn!

Finally, to my best friend and dear, dear husband Ole: thank you for filling my life with love and music, and for always caring more about me than about my thesis. Here it is. I'm back.

Nesodden, December 2013
Kathrine Frey Frøslie

Contents

Articles

I. Friis CM*, Frøslie KF*, Røislien J, Voldner N, Godang K, Ueland T, Bollerslev J, Veierød MB, Henriksen T. *The interleukins IL-6 and IL-1Ra: a mediating role in the associations between BMI and birth weight?*
Journal of Developmental Origins of Health and Disease 2010;5:310-318.
* Joint first authors.

II. Frøslie KF, Godang K, Bollerslev J, Henriksen T, Røislien J, Veierød MB, Qvigstad E. *Correction of an unexpected increasing trend in glucose measurements during 7years recruitment to a cohort study*.
Clinical Biochemistry. 2011;44:1483-6.

III. Frøslie KF, Røislien J, Qvigstad E, Godang K, Voldner N, Bollerslev J, Henriksen T, Veierød MB. *Shape information from glucose curves: Functional data analysis compared with traditional summary measures*.
BMC Medical Research Methodology 2013;13:6

IV. Frøslie KF, Røislien J, Qvigstad E, Godang K, Bollerslev J, Henriksen T, Veierød MB. *Shape information in repeated glucose curves during pregnancy provided significant physiological information for neonatal outcomes*.
Under revision for resubmission to PLoS ONE.

The articles are referred to by their Roman numeral throughout the thesis.

1 Introduction

1.1 Background

Statistical methods are continuously developed and improved upon, but the application of this growing body of methodological knowledge in medical research is often slow. Modern challenges, like the analysis of large data sets or longitudinal data with complex correlation structures, continue to be met by traditional and simple statistical methods, or ad hoc amendments to these. While simplifications of data structures and application of traditional statistical methods may in some cases work well, such approaches may also fail to reveal interesting information, or worse, simplify the analysis to the extent that results no longer reflect the research question. Methods targeted to solve specific problems, including modern statistical methods, are needed when studying problems of physiological and clinical interest in data with complex structures.

To understand physiological mechanisms, it is often necessary to disentangle and specify pathways that link an exposure to an outcome. The effect might be direct, or indirect through an intermediate variable, a so-called mediator. A mediator is a variable on the causal pathway between the exposure and the outcome (1;2). In medical research, a common approach to analyse the role of possible mediators, so-called mediation analysis, is to use standard regression models to estimate the effect of the exposure without adjusting for the mediator, and compare it with the estimated effect after adjusting for the mediator. Given specific conditions this approach may be justified and work well in problems with a simple structure (2). However, such an approach does not generally hold (2). In the field of metabolism and foetal growth, mediation and intermediate mechanisms are often presented by even simpler or unclear analyses (3-5). Significant pathways thus remain undetermined. Mediation is also often discussed on a principal level, without formal statistical analysis or comparison of effect estimates (6-8).

Path analysis, a special case of structural equation modelling, is a formal way of approaching mediation analysis (9;10). In a path analysis, the first step is to draw a path diagram, often called a directed acyclic graph (DAG) (1). Based on this path diagram, regression equations are linked in order to estimate the composite or indirect effect of mediators. The use of structural equation modelling and path analysis is widespread in the social sciences, and is increasingly used in medical research (11-14). Also, the last decades'

developments in the field of causality have increased the attention to causal pathways and DAGs (15-17).

Metabolic regulation is comprised of inherently continuous and temporal processes in the body. Some processes develop relatively slowly, like the increasing insulin resistance and low-grade systemic inflammation that follows from obesity (18). Other processes are fast, like the pulsatility of insulin secretion (19). While such temporal information is characteristic of the process, temporal information from metabolic processes is often discarded in clinical studies, and analyses are based on measurements of features that are relatively stable over time, like body mass index (BMI), or measurements from a stable part of the dynamic process, like fasting glucose or fasting insulin (20). Studies that incorporate temporal features often express such information by simple summary measures, like area under the curve (AUC) (21), a difference between two curve points, or the maximum amplitude (22).

However, temporal measurements can hold important information that is not revealed by such traditional summary measures. Functional data analysis (FDA) is a collection of statistical techniques specifically developed to analyse curve data, for example the result of temporal measurements of an underlying continuous process (23-26). When applying FDA, the entire curve is used as the basic unit of information, instead of single measurements at specific time points, or simple summary measures. FDA has given novel insights of clinical importance in several research areas (27-31), and a recent review advocates wider application of FDA (32).

1.2 Obesity

The prevalence of overweight and obesity is often assessed by BMI, defined as the weight in kg divided by the square of the height in meters (kg/m^2) (33). BMI is accepted as an indicator of body fat also in pregnant women, particularly in early pregnancy (34;35). The World Health Organization (WHO) classification for BMI is underweight (<18.5 kg/m^2), normal weight (18.5-25 kg/m^2), overweight (25-30 kg/m^2) and obese (≥ 30 kg/m^2) (33). These cut-off values are common benchmarks for assessment of prevalence of obesity, but the risks of disease or adverse pregnancy outcomes can increase from BMI values lower than the cut-off values for obesity or overweight.

Obesity has reached epidemic proportions globally (33;36), and obesity rates have risen also among pregnant women (37-39). The rising epidemic reflects the profound changes

in society and in behavioural patterns of communities over recent decades, especially the increased consumption of energy-dense foods and reduced physical activity (36).

Obesity implies profound metabolic changes that cause insulin resistance, high blood pressure and high levels of cholesterol and triglycerides, and is a major risk factor for serious diseases in the non-pregnant population, including diabetes mellitus (DM), cardiovascular diseases and certain forms of cancer (33;40). The adverse effects of obesity, particularly increased insulin resistance, are also observed in pregnancy.

Obesity in pregnancy is associated with increased risks of short-term and long-term outcomes for the woman. The short term ones include gestational diabetes mellitus (GDM), preeclampsia and caesarean section. In long term, obese women have increased risk of subsequent type 2 DM (34;39;41;42). Obesity is also an established risk factor for a long list of adverse outcomes for the foetus, including overgrowth, macrosomia, unfavourable neonatal body composition, and neonatal hypoglycaemia (37;39;41;42). Macrosomic neonates further have increased risk for obesity in adulthood (37), and a vicious circle of obesity across generations have been found (43). Children of obese mothers have also been found to have increased risk of cardiovascular diseases (44).

Obesity and inflammation

One of the well-documented effects of obesity is a systemic low-grade inflammatory state (18;45). This inflammatory state has been postulated to play a role in the progression of insulin resistance, DM and coronary heart diseases (18). The elevated levels of inflammatory markers in obese, pregnant women have also been suggested to play a role in the development of GDM (46) and adverse pregnancy outcomes, such as preeclampsia (11). Inflammation might also affect foetal growth, partly by modifying the glucose levels through increased insulin resistance (46;47), and partly by other processes, like affecting the placental transfer (48). The role of inflammation in the association between BMI and birth weight needs further study.

The research field of inflammation is large, and selecting the appropriate inflammatory markers is challenging. The adipose tissue has been found to secrete a large variety of proteins, including cytokines. Many of these have both immune-modulatory functions and act as systemic or auto-/paracrine regulators of metabolism, and may provide a mechanistic link between obesity and the associated complications (49). A wide range of cytokines can be considered, e.g. the interleukines IL-1Ra, IL-6, IL-10, monocyte

chemoattractant protein-1 (MCP-1), C-reactive protein (CRP) and the tumor necrosis factor TNF- RII (49). The interleukins IL-1Ra and IL-6 are known to be pro-inflammatory cytokines and are produced by both the placenta and adipose tissue. In addition, these markers are associated with obesity and their metabolic effects have gained extensive attention (45;50). In the present thesis, it was considered important to avoid a too complex path diagram, and the focus was restricted to the interleukins IL-1Ra and IL-6.

1.3 Glucose metabolism during pregnancy and GDM

During pregnancy, the mother substantially adapts her metabolism of carbohydrates, amino acids, lipids and vitamins to ensure supply of nutrients to the foetus and to meet the placental and maternal demands of late gestation and lactation (51;52). In normal pregnancy, there is a decrease in fasting glucose levels during the first trimester, before gestational week 12 (53). No consistent results have been found regarding a further decrease (or increase) in fasting glucose during the second or third trimester (53). Deterioration of glucose tolerance, i.e. insulin resistance, occurs normally during pregnancy, especially in the third trimester (52;54). The mechanisms of increased insulin resistance and secretion in pregnancy have been subject to many studies, and can be partially related to the metabolic effects of several placentally derived hormones and cytokines that are elevated in the maternal circulation during pregnancy (6;52).

Glucose measurements

There are many different approaches to measuring blood glucose, both in terms of biochemical equipment, what substance to measure, and how to conduct blood samples.

Biochemical analyses of blood samples are preferably done at a certified laboratory, but can also be done with simpler point-of-care devices (55). In either case, structured quality control should be performed to avoid analytical artefacts like time-dependent trends caused by the equipment, and to ensure comparability of the measurements (56).

Fasting glucose reflects the blood glucose level at glucose homeostasis after many hours of fasting (20), whereas glycosylated haemoglobin (HbA1c) reflects the average blood sugar over the past 1-3 months (57). Fasting glucose and HbA1c require a single blood sample only, are simple to make, and are used both diagnostically and as risk factors for a variety of disorders. However, neither captures the regulation of blood glucose after food

intake. The intravenous clamp technique, oral glucose tolerance tests (OGTTs) and continuous glucose monitoring provides data about glucose regulation.

The intravenous clamp procedures are considered to be the "gold-standard" in assessment of insulin resistance (58), but they are time-consuming, invasive and labour intensive. Although the use of OGTTs is debated (59), it is the simplest and most frequently used test procedure for large groups of people. In an OGTT, the response to orally consumed glucose is measured by repeated blood samples. Before the test begins, fasting glucose is measured. The amount of sugar (e.g. 50g, 75g, 100g), frequency of blood sampling (e.g. every 30 minute, every hour), and length of the test (1-3 hours) vary between different OGTTs. Other substances, e.g. insulin, may be measured in addition to blood glucose. Glucose measurements from OGTTs in pregnant women have shown elevated postprandial values and a delayed postprandial peak with increasing gestational age (60-64).

Categorisation of glucose measurements

Blood glucose regulation consists of continuous processes. However, few attempts have been made to analyse glucose curve characteristics formally (65-68). Clinical classification and diagnostic criteria are based on discrete glucose measurements (e.g. the fasting, 1-h and 2-h values), and categorization of these into terms like DM, impaired glucose tolerance, impaired fasting glycaemia or GDM (69).

GDM criteria prior to the Hyperglycaemia and Adverse Pregnancy Outcome (HAPO) study

GDM is defined as carbohydrate intolerance resulting in hyperglycaemia of variable severity with onset or first recognition during pregnancy (69). Several GDM criteria have been formulated. The pioneering O'Sullivan and Mahan GDM criteria from 1964 (70) were based on statistical identification of the upper limits of glycaemic normality during pregnancy (mean plus 2 SD). The cut-off values for fasting glucose, 1-h, 2-h and 3-h values after a 100 g OGTT were validated for their predictive value for subsequent development of type 2 DM in the mother (70-72). These criteria became the recommendations from the American Diabetes Association and the standard in medical care in the North America for almost 40 years (54;72;73).

Many other countries used the GDM criteria recommended by the WHO, which were based on cut-off values for fasting glucose and the 2-h value from a 75g OGTT (74;75). These cut-off values were updated in 1999, based on cut-off values for DM and impaired

Table 1.1. Most commonly used guidelines for the diagnosis of GDM. Table from World Health Organization (74).

Organisation	Fasting Plasma glucose	Glucose Challenge	1-h plasma glucose	2-h plasma glucose	3-h plasma glucose
WHO 1999*	≥ 7.0	75g OGTT	Not required	≥ 7.8	Not required
American Congress of Obstetricians and Gynecologists**	≥ 5.3	100g OGTT	≥ 10.0	≥ 8.6	≥ 7.8
Canadian Diabetes Association***	≥ 5.3	75g OGTT	≥ 10.6	≥ 8.9	Not required
IADPSG****	≥ 5.1	75g OGTT	≥ 10.0	≥ 8.5	Not required

* one value is sufficient for diagnosis
** two or more values are required for diagnosis
*** two or more values required for diagnosis
**** one value is sufficient for diagnosis

glucose tolerance in a non-pregnant population (Table 1.1) (54;69). Some institutions used only the cut-off for the 2-h value to define GDM, i.e. a single glucose measurement, as they considered the cut-off for the fasting value to be too high (74). Despite obvious disadvantages of having different GDM definitions across countries, no international consensus was found, and this motivated the HAPO study.

Maternal glucose and neonatal size: The Pedersen hypothesis, the HAPO study and new GDM criteria

Comparison of foetal growth and size of new-borns of pregnant women with DM and healthy pregnant women, led to the Pedersen hypothesis (76); that abundant glucose exposure for offspring of mothers with DM caused excess foetal insulin production which was the key in promoting foetal overgrowth and large neonates (Figure 1.1). Exposure to maternal DM was later also found to be associated with long-term effects for the offspring, e.g. type 2 DM and obesity (77).

The HAPO study (78) sought to clarify the risk of adverse pregnancy outcomes associated with glucose intolerance less than DM during pregnancy (78;79), and the study was expected to provide sufficient data as a basis for a more rational definition of GDM (80). The study sample of 25,505 pregnant women was impressive. Glucose tolerance was

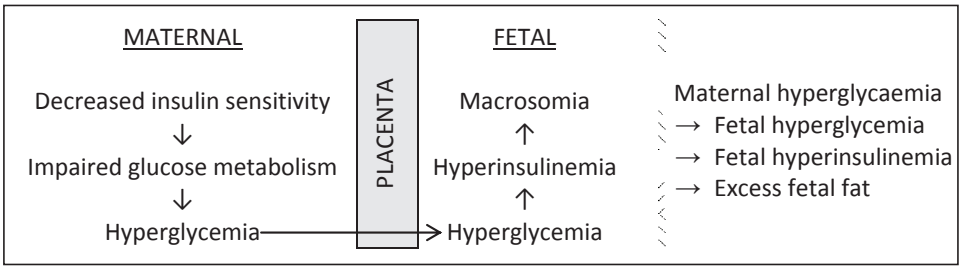


Figure 1.1. The Pedersen hypothesis

measured by a 75-g 2-h OGTT at 28 weeks of gestation, and the participating women, caregivers and HAPO staff were blinded to glucose tolerance values, except when predefined thresholds were met. The first findings were linear increases in risk of several pregnancy outcomes (e.g. macrosomia, cord C-peptide and neonatal percentage body fat) for increasing glucose levels below thresholds for DM (79;81). The HAPO study thus extended the Pedersen hypothesis: Maternal glycaemia, also in the normal range, affect the foetus (51). The continuously increasing risks, with no obvious thresholds at which risk increased more rapidly, were found for the fasting, 1-h and 2-h glucose values. Therefore, a consensus between clinicians and researchers was required to translate the results into clinical practice and a new GDM diagnosis.

In 2010, the International Association of the Diabetes and Pregnancy Study Groups (IADPSG) suggested that the new GDM criteria should be based on the fasting glucose and the 1-h and 2-h values from a 75 g OGTT (82). New cut-off values were based on the risk of adverse pregnancy outcomes (for mother and child), rather than cut-off values for non-pregnant people, or women's long-term risk of developing DM. Due to the linear increase in the risk found in the HAPO study, the cut-off values were chosen so that the odds ratio (OR) for the outcomes at the specified values were 1.75, relative to the mean glucose values. It was also decided that it was sufficient to exceed one of the cut-off values to be classified with GDM. The IADPSG criteria (Table 1.1) became the recommendations of the American Diabetes Association from 2011 (83), and the WHO from 2013 (74).

The prevalence of GDM in different populations around the world, using the criteria prior to the IADPSG, varies from 0.6% to 15% (51;84;85). It is generally accepted that the prevalence of GDM is higher with the IADPSG criteria. For instance, the HAPO researchers estimated that 17.8% of the HAPO participants would be classified with GDM with the

IADPSG criteria (80). Critics point to a practically inacceptable high prevalence of GDM and resources being spent on the wrong target group.

Did HAPO give the final answer?

Although recommended by influential institutions, the IADPSG criteria are subject to debate and controversy (80;86-88), and the American Congress of Obstetricians and Gynecologists is one of the organizations that do not recommend their use (Table 1.1). However, all present GDM criteria are based on three or more OGTT glucose measurements, in contrast to the WHO 1999 criteria, which were based on two, and in some institutions only one, glucose measurement. This implies that with the IADPSG criteria, the OGTT curve trajectory indirectly has gained attention, compared with the WHO 1999 criteria.

In summary, the HAPO study has provided essential knowledge about glucose metabolism in pregnancy, but has also given rise to new questions that need to be approached. There is still need for more knowledge that can enhance the understanding of the mechanisms involved in pathophysiology and adverse outcomes.

1.4 Birth weight and neonatal body composition

Birth weight and neonatal body composition are assumed to reflect foetal growth and the intrauterine environment, and are therefore important pregnancy outcomes (79;81;89-91). Excessive birth weight is associated with maternal birth complications, e.g. caesarean section, and shoulder dystocia. Prediction of birth weight and intervention to reduce risk factors for high birth weight are therefore the topics of many studies (92).

Birth weight is measured directly after birth and is often compared to the expected birth weight at a given gestational age, and transformed into a z-score, or categorised, e.g. into small for gestational age (SGA), large for gestational age (LGA) or macrosomia (79;89). Birth weight is a crude measure of neonatal body composition, but more information is obtained by skinfold caliper measurements or dual-energy X-ray absorptiometry (DXA) (81;93).

In Norway, the mean birth weight and proportion of macrosomic new-borns (defined as birth weight above 4000g or 4500g) have increased during the late decades of the 20th century, reaching a peak around the year 2000 (Figure 1.2) (94). Similar observations have been reported in several countries worldwide (95-98).

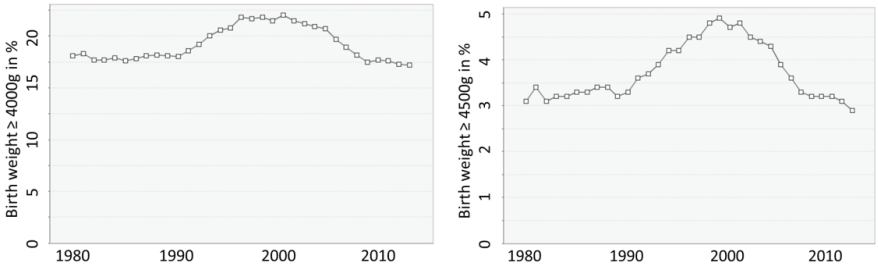


Figure 1.2. Percentage of macrosomic new-borns in Norway1980-2012. Source: Medical Birth Registry of Norway (94).

Birth weight has been widely used to predict future health outcomes, in particular coronary heart disease, overweight and DM (99-101), and has also been associated with brain development (102). Forsdal was the first to notice high rates of adult heart disease in areas of poor early life environment (103). Barker later found significant associations between low birth weight and atherosclerotic cardiovascular disease (CVD), hypertension, type 2 DM and insulin resistance, and this gave rise to the Barker hypothesis, also called the thrifty phenotype hypothesis (Figure 1.3) (103-106). The increased risk of adult disease for offspring with a poor intrauterine environment is increased further if those born small show rapid weight gain in childhood or become obese. The hypothesis is that CVD is "programmed" by under-nutrition during critical periods of early development and that a poor early life environment creates a permanent vulnerability to these diseases (104;107). It is assumed that there are different developmental windows for the programming of the hypothalamus, in which exposures like diet, physical activity and an impaired glucose regulation are more important (108). The concept of foetal programming through epigenetic mechanisms is the cornerstone in the concept of Developmental Origins of Health and Disease (DOHaD), and is not limited to effects of low birth weight (107;109).

Maternal BMI and glucose are modifiable variables that have consistently been found to affect neonatal size, i.e. birth weight and neonatal body composition (39;41;79;81;90;110), and are pointed out as targets for prevention of foetal overgrowth (90). The effect of maternal BMI on neonatal size can be divided into genetic factors, and factors influencing the intrauterine environment, such as placental function and nutritional supply to the foetus (111). The mediating role of inflammation has yet to be elucidated.

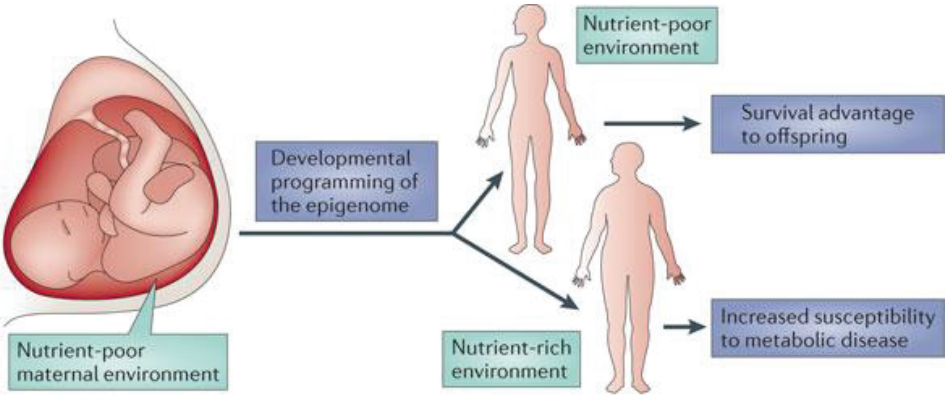


Figure 1.3 The thrifty phenotype hypothesis. From Walker and Ho (112).

Glucose is the primary energy source for the foetus, and the placental transport of glucose is higher than any other substrate (51). The alterations in the maternal metabolism in obesity, particularly the increased insulin resistance, cause elevated glucose levels in the maternal circulation. The placental transport of glucose is dependent of the maternal-foetal gradient across the placenta (113), and the extended Pedersen hypothesis (Figure 1.1) postulate that maternal glycaemia affects foetal growth mainly through increased insulin production in the foetus, thereby contributing to foetal hyperinsulinemia, as reflected in high levels of neonatal C-peptide, and various aspects of diabetic foetopathy, including foetal overgrowth, large neonates and deposition of body fat (81). The HAPO researchers noticed that no single glucose measurement was clearly superior to the others in terms of associations with outcomes. This was based on analyses of single glucose measurements from a single OGTT in gestational week 28. Longitudinal OGTT data from the pregnancy was not available, and no analyses explored the potential information in entire OGTT curves. The intense debate concerning the new GDM criteria is a powerful argument for searching for new information inherent in entire glucose curves, instead of single glucose measurements, also when exploring the effect on neonatal outcomes.

2 Aims of the thesis

The overall aim was to use path analysis to analyse mediation, and FDA to analyse OGTT data, to extract more information from a large cohort of pregnant women than can be done with simple, commonly used methods, and show that this gives more insight into physiological and clinical problems under study. An important additional aim was to facilitate the description of the methods and the presentation of the results, to illustrate the advantages and make the methods available to a wider audience.

More specifically, the aims were

- To study whether the interleukins IL-6 and IL-1Ra have mediating roles in the association between early pregnancy BMI and birth weight, using path analysis.
- To investigate whether an unexpected trend in the glucose measurements during the inclusion period was of biological or analytical origin, and, if not biological, to remove the trend by regression analysis.
- To study the usefulness of FDA in the analysis of OGTT glucose curves from one time point in pregnancy by comparing shape information extracted by FDA with standard simple summary measures. Furthermore, to analyse the shape information inherent in OGTT glucose curves from early pregnancy in relation to early pregnancy BMI categories and GDM later in pregnancy.
- To use multilevel FDA to study the shape inherent in OGTT glucose curves from two visits during pregnancy and to analyse the effect of such information on the neonatal outcomes birth weight, percentage fat and C-peptide in cord blood.

3 Material and Methods

3.1 The STORK study

Participants

The STORK ("STORe barn og Komplikasjoner"; "Big babies and complications") study is a prospective cohort of 1031 healthy women of Scandinavian heritage who registered for obstetric care at Oslo University Hospital Rikshospitalet from 2001 to 2008. Exclusion criteria were multiple pregnancies, known pre-gestational type 1 or type 2 DM, and severe chronic diseases (pulmonary, cardiac, gastrointestinal or renal). The overall aim of the study was to gain insights into maternal metabolic syndrome and determinants of foetal macrosomia (114).

The data collection was done at the Division of Obstetrics and Gynaecology and the Section of Specialised Endocrinology. Due to funding and logistics, two teams of investigators did the data collection at the obstetrics unit. Data from the first 553 women recruited to STORK during 2001-2005 formed basis for the first publications from the study (110;114-116). Data from the following 478 women were collected during 2005-2008 by the same routines. A bio-bank from the study contains frozen blood samples from the cohort for future use.

Data collection

Data were collected at five time points: at gestational weeks 14-16 (inclusion), 22-24, 30-32, 36-38 and birth. Data from inclusion, weeks 30-32 and birth were used in this thesis.

Gestational age at inclusion was estimated according to Naegele's rule (117). Routine ultrasound data at gestational weeks 17-19 were used to estimate gestational age at subsequent visits and at birth.

Blood samples at all visits during pregnancy were drawn in the morning, between 0730 and 0830 after an overnight fast, and were obtained from veni-puncture in tubes containing Ethylenediaminetetraacetic acid (EDTA). Blood samples were immediately put on ice, plasma isolated and stored at -80°C until analysed.

Age, height, parity, education and smoking and were registered at inclusion. Data on preeclampsia and hypertension were obtained from hospital records.

BMI (articles I-IV)

Maternal weight was measured at each visit. Early pregnancy BMI was calculated as weight at inclusion (kg), divided by self-reported height (m) squared.

Glucose (articles I-IV)

Fasting glucose and results from a 75 g OGTT were recorded at gestational weeks 14-16 and 30-32. Plasma glucose was measured immediately in a drop of fresh, whole EDTA blood. During the OGTT, blood samples were taken every 30 minute for two hours. Glucose measurements during the study period were done primarily by the Accu-Chek Sensor (ACS) glucometer (Roche Diagnostics GmbH, Mannheim, Germany). Inter-assay coefficient of variation was <10%.

Three different ACS electronic metres were used during the study period. Independent ACS quality control solutions in the low (hypoglycaemic) and high (hyperglycaemic) range were used and registered every time a new vial of glucose strips was opened, approximately once a week during the study period.

After the end of the inclusion period, we unexpectedly detected an increasing trend in the fasting glucose levels at inclusion, and in the ACS quality controls in the low and high range. To investigate this further, randomly sampled glucose values from the whole study period (2001-2008) were analysed by the hexokinase method at an accredited laboratory at Oslo University Hospital in 2011. The hexokinase data (n=170) were based on frozen serum from the 90 min OGTT at gestational weeks 30-32 and analysed by a Hitachi Modular P chemistry analyser with reagents from Roche. As a consequence of these and other investigations, all glucose measurements were de-trended prior to the analyses in articles III and IV, as described in detail in article II.

Classification of GDM in article III was based on the WHO recommendations at the time of publication, with a cut-off value for the 2-h value of 7.8 mmol/l.

Insulin (articles II and IV)

Measurements of insulin were obtained for the whole cohort from frozen blood serum at the end of the recruitment period (2007-2008). The insulin samples were assayed in duplicate by Radio Immuno Assay (Siemens Medical Solutions Diagnostics, CA, USA), and the means of the two measurements were used in analyses.

Inflammatory markers (article I)

IL-6 (high sensitivity) and IL-l Ra were measured by ELISA using commercially available kits (BIOSOURCE, Invitrogen Corporation). CRP was measured as described by Wu et al. (119). All samples were measured in duplicate and serial samples from a given individual were analysed at the same time to minimize the run-to-run variability. Intra-and inter-assay coefficients of variation were < 10% for all assays.

Neonatal outcomes (articles I and IV)

Birth weight (g) was measured with a calibrated scale within two hours after the birth. Umbilical cord blood was collected into EDTA tubes by the midwife, centrifuged for plasma separation and placed at -20 ºC for less than a month and stored long term at -80 ºC.

The percentage of neonatal body fat was measured by DXA scanning within 4 days postpartum, as described in detail in (93). Briefly, the neonates were scanned for approximately 6 minutes during sleep. DXA was primarily developed for the assessment of bone mass, but it also provides information on total fat mass and fat-free mass, as well as the tissue distribution in the trunk and extremities. Lunar Prodigy software (version 12.10) was developed especially for infant DXA and was used to analyse all scans (93).

Plasma levels of C-peptide in cord blood samples were measured using a RIA from Millipore (Corporation, Billerica, MA, USA) (120). Assays were performed according to the manufacturer's instructions. The intra- and inter-assay coefficients of variation were < 10% for all assays.

3.2 Study samples

Article I

Inflammatory markers were obtained from 240 women from the first part of the study, due to the limited resources for biochemical analyses. The subsample was based on women enrolled in the cohort during 2001-2005 and restricted to those who gave birth to a baby with a birth weight above the 10th birth weight percentile of the cohort. Stratified random sampling based on birth weight below or above 4200 g was used to ensure that women with macrosomic babies were included in the subsample. Women with possible infections at the time of sampling, indicated by a CRP value above 10 mg/l, extreme values on IL-l Ra or IL-6, or missing data of any variable in the analyses were excluded, and the final study sample consisted of 208 women (Figure 3.1).

Article II

The entire STORK cohort was used in article II.

Article III

The 974 (94%) women with complete OGTT data at inclusion in gestational weeks 14-16 constituted the study sample for the FDA (left column in Figure 3.2). Regression analyses were restricted to women with complete data of all variables in the analyses: 966 women in the analysis of early pregnancy BMI, and 922 women in the analysis of the categorised 2-h glucose value at weeks 30-32.

Article IV

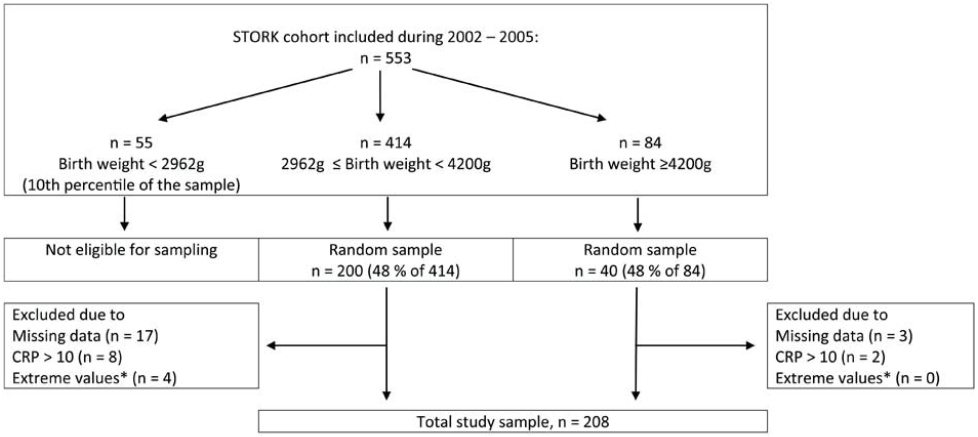
The 884 (86%) women with complete OGTT data at gestational weeks 14-16 and 30-32 constituted the study sample for the FDA (right column in Figure 3.2). DXA scanning data and cord blood C-peptide were obtained for 207 neonates from the last part of the STORK study. Regression analyses were restricted to women with complete data of all variables in the analyses of birth weight (n=868), neonatal percentage fat (n=185) and cord blood C-peptide (n=134).

3.3 Ethical considerations

The study was approved by the Regional Committee for Medical Research Ethics, Southern Norway, Oslo, Norway (reference number S-01191), and performed according to the Declaration of Helsinki. All participating women provided written informed consent.

3.4 Statistical analyses

The main method in article I was Bayesian path analysis. The analyses of changes over time in article II included use of statistical process control and linear and local linear regression models. Articles III and IV applied various FDA methods. Functional principal component analysis (FPCA), functional analysis of variance (FANOVA) and traditional nominal logistic regression analysis were used in article III. The main method in article IV was multilevel Bayesian FDA, including FPCA, as well as traditional linear regression analysis. The analyses are described in more detail below.



*Women with extreme IL-1Ra values or IL-6 values, that is, values outside ±3SD in log scale, were excluded from the analysis. This corresponded to IL-1Ra values below 48 pg/ml or above 626 pg/ml, and IL-6 values below 0.02 pg/ml or above 1.88 pg/ml.

Figure 3.1. Flow-chart for article I.

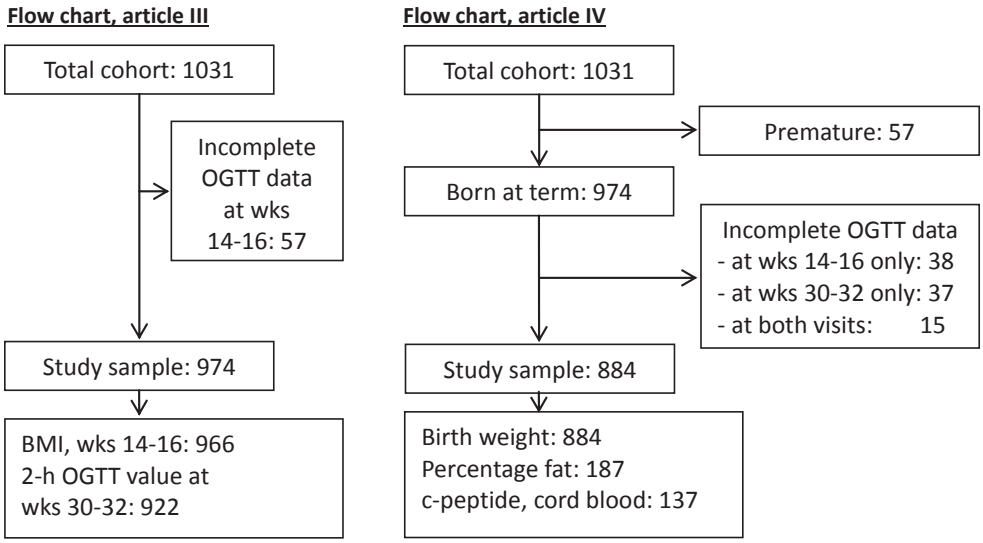


Figure 3.2. Flow charts for articles III and IV.

Data description (articles I-IV)

Descriptive statistics were presented as mean, standard deviation (SD) and range, or frequency and percentage (%), or as median and quartiles. In article I, independent samples *t*-tests were used to compare the study sample (n=208) and the remaining eligible women in the STORK cohort included during the same time period (n=258). The study samples in articles III and IV, and women with incomplete OGTT data were compared by two-sample *t* tests or χ^2 tests where applicable.

Simple summary measures of OGTT (articles III and IV)

In article III, simple summary measures of OGTT included the fasting value, 2-h value, AUC (trapezoidal rule) and the most cited simple shape index, defined by Tschritter et al. (66). This index is defined as the 2-h value minus the 90-min value for curves classified as "monophasic" or "biphasic", and the 90-min value minus the 60-min value for curves classified as "triphasic". The classification of curves, i.e. the determination of the number of phases within a curve is based on an empirically chosen glucose threshold of 0.25 mmol/l (66). Curves that did not meet the criteria for classification into either mono-, bi- or triphasic were labelled "unclassified" and left out of the analyses. In article IV, only AUC was used in addition to FDA.

Categorisation of variables (article III)

Early pregnancy BMI was categorised according to the WHO classification (underweight (<18.5 kg/m^2), normal weight (18.5-25 kg/m^2), overweight (25-30 kg/m^2) and obese (≥30 kg/m^2)) (33). As described in section 1.3, the discussion about the GDM criteria was not settled in WHO when article III was published, but the 2-h OGTT cut-off of 7.8 mmol/l (Table 1.1) was important in clinical practice. To visually demonstrate the clinical usefulness of the curve shape information, the 2-h values at weeks 30-32 were grouped into seven categories and used as the outcome in analyses of curve shape information. The 2-h value categories were based on the diagnostic criterion for GDM and on assessments of group size and percentiles in the sample: <3.27 (2.5th percentile), [3.27, 3.89) (2.5th-10th percentile), [3.89, 6.39) (10th-75th percentile; reference category), [6.39, 6.90) (75th-85th percentile), [6.90, 7.8) (85th percentile to diagnostic cut-off for GDM) [7.8, 8.84) (GDM diagnosis to 98th percentile) and ≥8.84 mmol/l.

3.4.1 Path analysis (article I)

Path analysis is a multivariable statistical methodology, linking a series of regression equations together (9), with the aim of estimating the composite or indirect effect of both explanatory variables and mediators. Within this system of equations, some of the variables can be considered both as outcome variables and as explanatory variables. Path analysis is a special case of structural equation modelling, in which all hypothesized dependencies between the variables are specified in a model and depicted in a path diagram prior to the analysis. The arrows in the path diagram represent dependencies between variables, and absence of an arrow between two variables indicates that these variables are considered to be conditionally independent (121). The effect of an exposure that acts on the outcome through a mediator is termed an *indirect* effect, while the effect of an exposure on the outcome that is not explained by a mediator is termed a *direct* effect. All indirect and direct relations among measured variables can be read off the diagram.

Based on the literature we constructed a path diagram, including early pregnancy BMI, birth weight, and the inflammatory markers IL-6 and IL-1Ra and fasting glucose in weeks 30-32 (Figure 3.3). All variables were entered into the analysis as standardized variables in order to quantify the relative importance of factors within the study. Indirect effects were calculated by multiplication of the standardized regression coefficients along a given path, and total effects were found by summing all direct and indirect effects between two variables.

The path analysis was performed using Bayesian estimation procedures. The Bayesian analysis gives estimates of regression coefficients and corresponding credibility intervals (CrIs), which are comparable with frequentistic confidence intervals (CIs). Considerations of statistical significance can be based on the coverage of the CrIs. Comparison of path models was carried out by comparing values of the deviance information criterion (DIC); models with lower values of DIC are preferable.

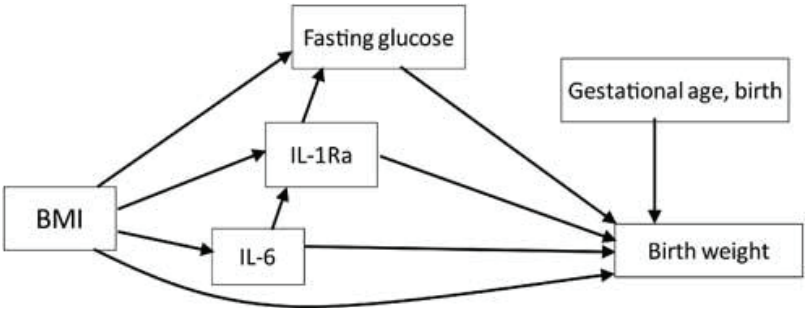


Figure 3.3: The path model
Path diagram showing a decomposition of the hypothesized effect of early pregnancy BMI on birth weight. The indirect pathways between BMI and birth weight were hypothesized to be mediated by fasting glucose (nutrient availability), the interleukins IL-1Ra or IL-6 or a combination of these. BMI was measured at inclusion, fasting glucose and interleukins at weeks 30–32 and gestational age at birth. Arrows represent dependencies between variables. Absence of an arrow between two variables indicates that the variables are considered to be statistically independent in the model.

3.4.2 Trend in time series data (article II)

Processes measured repeatedly over time generate time series data. A long-term movement in such a time series is called a trend (1), and an essential feature of a trend is a consistent change over the whole time interval under study. Statistical process control (SPC) procedures provide useful tools such as control charts, to monitor process behaviour over time, and to detect and prevent errors or bias in measurement procedures (122). The control chart called "X-chart" displays the time series data as well as the mean and so-called control limits, often also alarm limits, based on the variability in the data, i.e. 3 SDs and 2 SDs, respectively (122). A variety of rules have been developed to detect whether a process is stable or changing, both mathematical and more pragmatic ones. For simple practical use, detection rules are often reduced to a few rules-of-thumb (122). The simplest and most common rules are

- A single data point outside a control limit
- Two out of three successive values outside the same alarm limit
- Eight or more successive values on the same side of the mean
- Six or more values in a row steadily increasing or decreasing.

Using regression to de-trend time series

De-trending refers to the mathematical operation of removing a trend from a time series (123), and is often applied to remove a feature assumed to distort or obscure the measurements or relationships of interest. A linear trend in the mean can be removed by subtracting a least-squares-fit straight line from a linear regression analysis. Less straightforward trends might require more advanced procedures. Local regression lines (124) can be an alternative when nothing is known about the underlying reasons for the change over time. De-trending procedures can be based on intrinsic structures in the data, or on an external sample assumed to validate the data, like independent control samples at a laboratory.

Time-dependent trends were explored by scatter plots and linear regression analyses, local polynomial regression analyses (124) and X-charts (56). Linear regression gave estimates for the average increase per time unit, under the assumption of a linear increase during the entire period. De-trended glucose values were estimated by a weighted average of the regression coefficients from the independent low and high control solution values (124), with weights chosen as the inverse distance from the measured to the predicted glucose value for the low and high controls, and scaled to sum to one. Thus, a woman's de-trended (adjusted) glucose level was expressed as

$$y_{i,adj} = y_i - \left(w_i \cdot \hat{\beta}_{low} + (1 - w_i) \cdot \hat{\beta}_{high}\right) \cdot t_i \,, \qquad (1)$$

where y_i is the observed fasting glucose for the *i*th woman, *i*=1, 2,..., *n*. The time from start of the study is t_i and $t_i \in [0,7]$ years. Further, $\hat{\beta}_{low}$ and $\hat{\beta}_{high}$ are the slope estimates from the linear regression analyses of glucose on time in the low and high controls, respectively. The weights are given by

$$w_i = \frac{\dfrac{1}{\left|y_i - \hat{y}_{i,low}\right|}}{\dfrac{1}{\left|y_i - \hat{y}_{i,low}\right|} + \dfrac{1}{\left|\hat{y}_{i,high} - y_i\right|}}$$

where $\hat{y}_{i,low}$ and $\hat{y}_{i,high}$ are the predicted values from the linear regression analyses in high and low controls.

3.4.3 Functional data analysis (FDA, articles III and IV)

In FDA the basic units of information is not a single data point, but entire curves, varying smoothly over a continuum. The continuum is often time, like in our applications. FDA makes it possible to extract information from a temporal process as a whole, instead of merely point-by-point. In a sample of curves, the mean curve is used descriptively, as in traditional statistical analyses, and with proper modifications, most standard statistical methods can be phrased in the framework of FDA.

Functional data may represent both long and short-term time processes, with OGTT glucose values as an example of the latter. If such data are collected repeatedly over time as in the STORK study, a multilevel functional model must be applied.

Fitting continuous and individually smoothed curves

The OGTT glucose measurements were converted into continuous, smoothed glucose curves by subject-specific spline smoothing with B-splines basis functions, as described by Ramsay and Silverman (Appendix A, articles III and IV) (23-25). These individually fitted curves formed the basis for the subsequent FDA. In article III, the five OGTT measurements from weeks 14-16 for the 974 women in the study sample resulted in 974 corresponding glucose curves. In article IV, the OGTT measurements for the 884 women at two visits were converted into 884 glucose curves from weeks 14-16, and 884 glucose curves from weeks 30-32.

Functional principal component analysis (FPCA)

FPCA was used to study the temporal variation in the fitted glucose curves. FPCA extracts functional principal component (FPC) curves that describe the main modes of temporal variation in the sample of glucose curves (Appendix B in article III and Appendix A in article IV) (24). The FPC curves represent independent parts of the overall variability between the glucose curves and are given in descending order according to the proportion of explained variance. The FPCA also yields individual FPC scores for each glucose curve. The score variables are constructed to be uncorrelated, and the variation within the scores of an FPC quantifies the magnitude of the total variance explained by this FPC. A woman's FPC score for an FPC curve reflects to what extent her individual curve trajectory corresponds to the general temporal feature expressed by this FPC curve. Applying FPCA it is thus possible to study how glucose curve trajectories vary between women. As in traditional principal

component analysis, FPCs may be interpreted and labelled according to the information they exhibit, which in turn can potentially be related to conventional physiological or clinical theories.

In article III, the FPC curves and FPC scores were estimated simultaneously (Appendix B). In article IV, we first used a multilevel functional model to decompose the curves into subject- and visit-specific contributions to the overall mean. We then used the corresponding covariance matrices to estimate the FPC curves, and finally a Bayesian analysis to estimate the FPC scores. This analysis is outlined below, and details are given in Appendix A in article IV.

The functional multilevel model and multilevel FPCA (article IV)

A multilevel model for functional data was used for the analysis of glucose curves from two visits (125;126). Assume that the individual, true blood glucose curve $\gamma_{iv}(t)$ for woman $i = 1, \cdots, 884$ at visit $v = 1, 2$ in the continuous time span from 0 to 120 minutes, $t \in [0, 120]$, can be decomposed into fixed and random effects curves (Figure 3.4A), and expressed as

$$\gamma_{iv}(t) = \mu(t) + \eta_v(t) + X_i(t) + U_{iv}(t). \qquad (1)$$

Here the fixed effects curves are the overall mean glucose curve $\mu(t)$ (Figure 3.4Ai), and the mean visit-specific deviation from the overall mean curve, $\eta_v(t)$ (Figure 3.4Aii). Together, these terms constitute the visit-specific mean curve, $\mu(t) + \eta_v(t)$. The random effects curves are $X_i(t)$, the subject-specific deviation from the visit-specific mean curve, and $U_{iv}(t)$, the subject- and visit-specific deviation from the subject-specific mean curve. The curves $X_i(t)$ in expression (1) are unknown until they are estimated by a linear combination of the first FPC curves of $X_i(t)$, $\psi_a^X(t)$, $a = 1, 2$ (Figure 3.4B), and corresponding, estimated individual score variables, giving $\hat{X}_i(t)$, $i = 1, \ldots, 884$ (Figure 3.4Ciii). Likewise, the curves $U_{iv}(t)$ in expression (1) are unknown until they are estimated by a linear combination of the first FPC curves of $U_{iv}(t)$, $\psi_b^U(t)$, $b = 1, 2, 3$ (Figure 3.4B), and corresponding, estimated individual score variables, giving $\hat{U}_{iv}(t)$, $i = 1, \ldots, 884$ and $v = 1, 2$ (Figure 3.4Civ). Details of the FPCA are given in Appendix A in article IV.

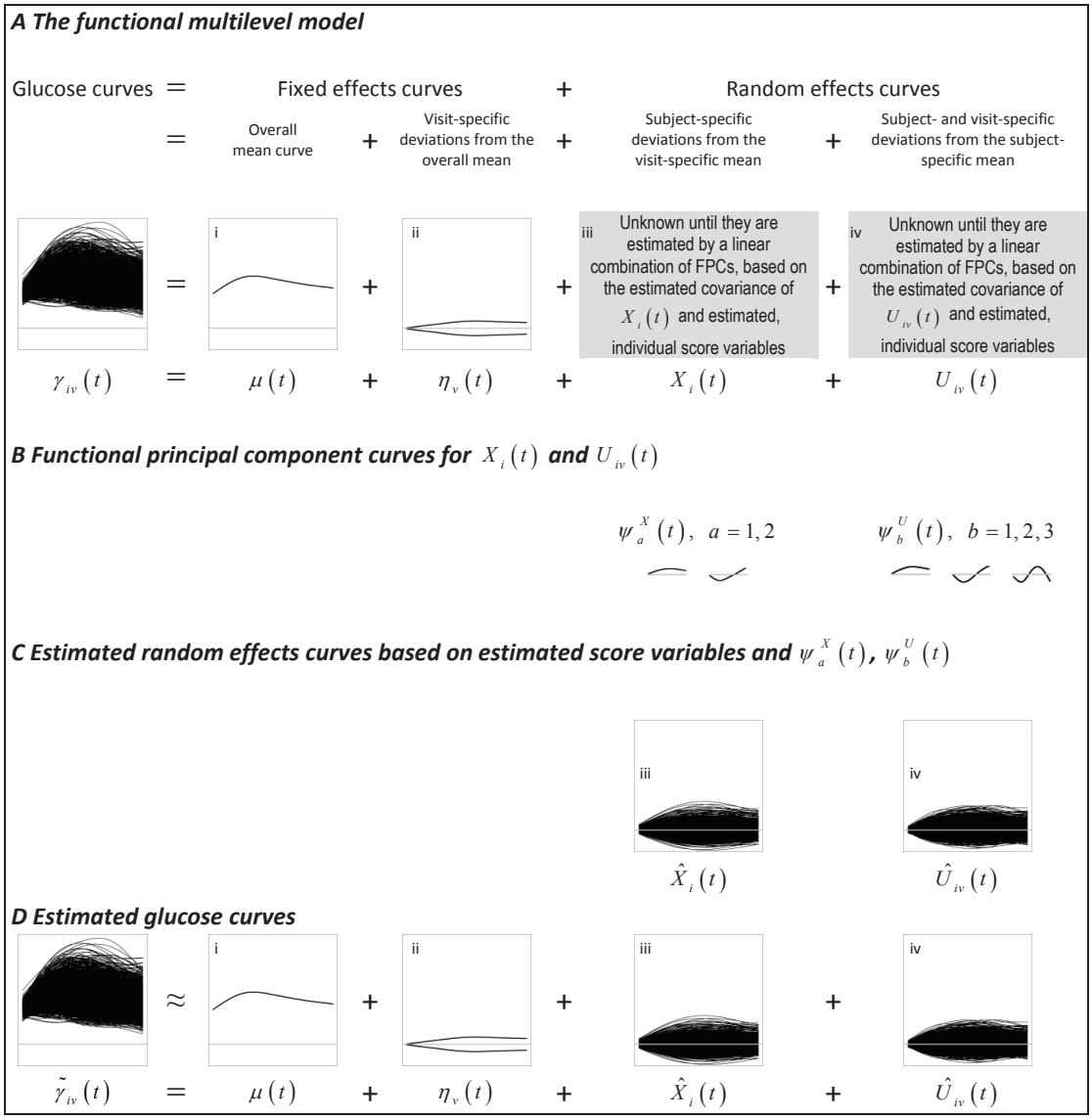


Figure 3.4. The functional multilevel model. In all plots, the horizontal axis is time during the 2-h OGTT, and the vertical axis is blood glucose, with range from -2 to 12.5 mmol/l. The horizontal, grey line is 0 mmol/l. $\gamma_{iv}(t)$ is the glucose curve from 0 to 120 min for woman $i = 1, \cdots, 884$ at visit $v = 1, 2$. $\hat{X}_i(t)$ and $\hat{U}_{iv}(t)$ are the estimates of $X_i(t)$ and $U_{iv}(t)$, using the first two subject-specific FPCs and the first three subject- and visit-specific FPCs, respectively. $\tilde{\gamma}_{iv}(t)$ is the estimated glucose curve, using $\hat{X}_i(t)$ and $\hat{U}_{iv}(t)$.

A woman's scores for the FPC curves of $X_i(t)$ quantify her subject-specific deviation from the visit-specific mean curve, i.e. the important characteristics of her glucose curves *across* visits (Figure 3.4Ciii and 3.4Diii). Her scores for the FPC curves of $U_{iv}(t)$ quantify her subject- and visit-specific deviation from her subject-specific mean curve, i.e. the characteristics of the residual variation *within* a visit (Figure 3.4Civ and 3.4Div).

By combining equation (1) with the FPC curves and corresponding estimated FPC scores, an individual glucose curve can be expressed as the sum of the visit-specific mean, $\mu(t)+\eta_v(t)$, and a linear combination of a small number of the FPC curves for $X_i(t)$ and $U_{iv}(t)$ (Figure 3.4D and equation (3) in Appendix A in article IV).

FDA vs simple summary measures

The Pearson correlation coefficient (*r*) was used to assess the associations between FPC scores, original glucose measurements and the simple summary measures of OGTT: Between fasting value, 2-h value, AUC and the shape index in article III, and between FPC scores, glucose measurements and AUCs in article IV. In article III, the simple summary measures were compared across categories of early pregnancy BMI using traditional ANOVA, with Bonferroni corrected post hoc tests. These results were then compared with the FANOVA results.

F*unctional analysis of variance (FANOVA) (article III)*

FANOVA, the functional counterpart of traditional analysis of variance (ANOVA), was used to compare the shape of glucose curves across categories of early pregnancy BMI (underweight, normal weight, overweight and obese). BMI was entered as an explanatory variable, and the fitted curves were the functional outcome. The analysis was based on the shape of the mean curve in each BMI category, and the temporal differences between these mean curves (Appendix C). Functional 95% CIs and *p* curves were obtained for the difference between two mean curves. Overall *p* values for the differences between two BMI categories can be obtained from the maximum value of the test-statistic used to compare curves, and these are also presented.

Curve shape information in regression analyses

In article III, the impact of the curve shape in early pregnancy on glucose intolerance later in pregnancy was assessed by regression analyses, using the FPC scores at weeks 14-16 as explanatory variables, and the categorised 2-h glucose value at weeks 30-32 as the outcome. Five different models were fitted. Model 1 included early pregnancy BMI and the three independent FPC score variables from weeks 14-16 as covariates, while models 2-5 included BMI and either the fasting value, the 2-h value, the AUC or the shape index, all from weeks 14-16, as covariates. These simple measures were included one at a time in models 2-5, due to colinearity. All covariates were continuous.

In article IV, the impact of glucose curve characteristics on the neonatal outcomes birth weight, percentage fat and C-peptide in cord blood were estimated using linear regression with FPC scores from the multilevel FPCA as explanatory variables. The interpretation of the effect estimates is based on the physiological interpretation of the FPC scores. Adjusted effect estimates were found by multiple linear regression analyses with most FPC scores (the first subject- and visit-specific score at weeks 14-16 was left out due to colinearity issues), early pregnancy BMI, age and parity as explanatory variables. The multivariable analyses involved stepwise variable selection procedures based on Akaike's information criterion, analyses of several models considered to be of importance, and considerations of physiological importance of the findings. The final multiple models include only the variables identified by these procedures. Model diagnostics were thoroughly checked during the analysis.

Software

SPSS 19 was used for the STORK data base, and for simple statistical analyses in all articles. The Bayesian analyses in articles I and IV were done in WinBUGS, using the R package `R2WinBUGS` which provides functions to call WinBUGS from R (127;128). The control charts for article II were made in Excel. All other analyses were done in R (129). The R program code, including the WinBUGS model files, is provided as supplementary material to articles I, III and IV.

4 Summary of results

Article I: The interleukins IL-6 and IL-1Ra: a mediating role in the associations between BMI and birth weight?

Means (SDs) for early pregnancy BMI and birth weight were 24.9 kg/m^2 (4.2) and 3748 g (454), respectively. The direct effects on birth weight of BMI and fasting glucose expressed by standardised regression coefficients (95% CrI) were 0.16 (0.00, 0.32) and 0.14 (0.01, 0.27), respectively. The direct effect of IL-1Ra on birth weight was not statistically significant (0.06 (-0.10,0.21)), but significant effects of BMI on IL-1Ra (0.61 (0.51, 0.72)), of IL-1Ra on fasting glucose (0.17 (0.01, 0.34)) and of fasting glucose on birth weight (0.14 (0.01, 0.27)) implied an indirect pathway from BMI via IL-1Ra on birth weight.

The total effect of BMI on birth weight was 0.24 (0.12, 0.36) (Figure 4.1: 0.16+(<0.001)+0.02+0.03+0.02). The estimated effect of BMI involving IL-1Ra was 0.05 (-0.05, 0.15), i.e. approximately 20% (0.05/0.24) of the effect of BMI on birth weight was mediated through IL-1Ra. For IL-6, only a negligible percentage was found (<0.001/0.24). The remaining 67% (0.16/0.24) of the BMI effect represent effects not explained by variables or structures in our model.

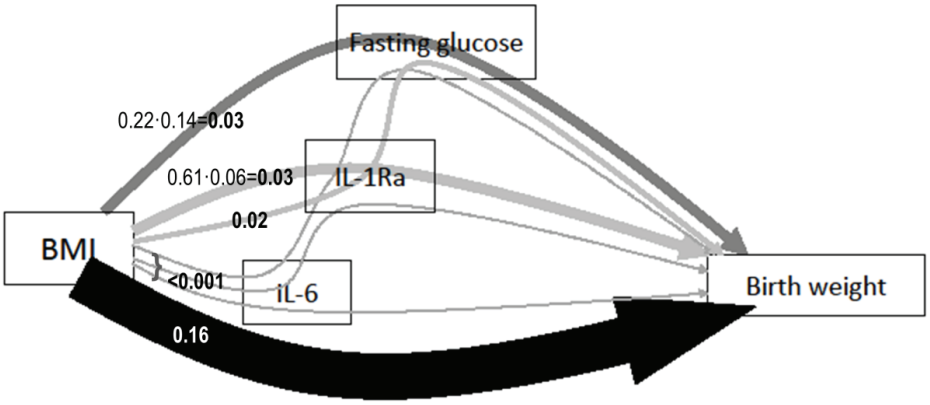


Figure 4.1. The figure shows the decomposition of the total effect of maternal BMI on birth weight. The total effect is the sum of all arrows, that is, the direct and indirect effects. The arrow widths represent the relative proportions of the total effect through a specific pathway.

Article II: Correction of an unexpected increasing trend in glucose measurements during 7years recruitment to a cohort study.

After the end of the inclusion period, we unexpectedly detected an increasing trend in the fasting glucose levels at inclusion. Mean (SD) fasting glucose measured by ACS increased from 4.0 (0.4) mmol/l for the first 100 women (2001-2002) to 4.6 (0.4) mmol/l for the last 100 women (2007-2008). The yearly trend was $\hat{\beta} = 0.11$, 95% CI (0.09, 0.12) mmol/l, p<0.001. Significantly increasing trends were also found for OGTT measurements at 30, 60, 90 and 120 minutes at inclusion in weeks 14-16, and 0, 30, 60 and 90 minutes at weeks 30-32 (Figure 1 in the Appendix), and for the low and high controls: $\hat{\beta}_{low} = 0.06$ (0.04, 0.08) mmol/l, p<0.001, and $\hat{\beta}_{high} = 0.49$ (0.42, 0.55) mmol/l, p<0.001, respectively. In contrast, the yearly trends for age, early pregnancy BMI and fasting insulin at inclusion were $\hat{\beta}_{age} = 0.09$ years (p=0.20), $\hat{\beta}_{BMI} = -0.17$ kg/m^2 (p=0.01) and $\hat{\beta}_{ins} = -0.08$ mmol/l (p=0.07), respectively, and local regression lines indicated weak negative curvatures in these trends. An overall non-significantly decreasing trend ($\hat{\beta}_{hex} = -0.12$ mmol/l, p=0.10) with negative curvature was also found for the 170 hexokinase data. The differences between the hexokinase data and the original 90 min OGTT measurements (weeks 30-32) increased significantly during the study period: $\hat{\beta}_{diff} = 0.24$ mmol/l (p<0.001). Accordingly, glucose measurements became increasingly biased upwards as time passed.

The shift in the low and high controls would have been detected during the inclusion period if continuous use of control charts with fixed mean and control limits based on the first 30 measurements had been applied (Figure 2 in the Appendix).

The increasing trend in the fasting glucose measurements at inclusion was successfully removed by subtracting a weighted average of the linear regression results from the independent control solutions.

The OGTT measurements at 30, 60, 90 and 120 minutes in weeks 14-16, and 0, 30, 60, 90 and 120 minutes in weeks 30-32 were also de-trended using the same technique (Figure 1 in the Appendix). This was not described in article II.

Article III: Shape information from glucose curves: Functional data analysis compared with traditional summary measures.

The smoothed glucose curves at gestational weeks 14-16 showed large variation between the individual curves. Over 99% of this variation was expressed by the first three FPCs, interpreted as "general level" (FPC1, 88.1%), "time to peak" (FPC2, 8.6%) and "oscillations" (FPC3, 2.4%).

The 2-h value was positively associated with all three FPC scores ($0.37 \leq r \leq 0.79$), in contrast to the fasting value ($-0.12 \leq r \leq 0.47$). AUC was highly correlated with the FPC1 scores ($r=0.999$) but not with the FPC2 and FPC3 scores ($r=-0.01$ and $r=0.05$, respectively). The shape index was most strongly associated with the FPC3 score ($r=0.67$), the principal component explaining the smallest part of the total variation.

The means of the glucose curves at weeks 14-16 differed between the four BMI categories in early pregnancy: While the curvature was similar, the levels of the mean curves for normal weight, overweight and obese women were significantly different ($p<0.001$). No significant difference was found between underweight and normal weight women ($p=0.26$).

The means of the fitted glucose curves at weeks 14-16 for the seven pre-defined categories of 2-h values at weeks 30-32 showed that the general glucose levels at weeks 14-16 were different in the 5 lowest categories, and that the mean curves in the two upper categories (from women diagnosed with GDM at weeks 30-32, n=51), displayed different pathophysiology at weeks 14-16. This was confirmed in multinomial logistic regression analyses with the seven categories of 2-h values at weeks 30-32 as the outcome: The FPC1 scores and the AUC yielded nearly identical results and were significantly different in the five lowest groups ($p \leq 0.02$), whereas no significant difference was found between the two subgroups with GDM ($p=0.40$), or between the two closest GDM and non-GDM groups ($p=0.59$). Also, no significant differences were found for the fasting value, 2-h value or shape index in the three upper categories ($0.07 \leq p \leq 0.92$), i.e. between subgroups of women with and without GDM. In contrast, mean FPC2 scores were significantly different between women who did and did not develop GDM, and between subgroups of women diagnosed with GDM later in pregnancy ($p=0.01$ and $p=0.02$), respectively. Also, FPC3 scores between the two GDM categories were significantly different ($p=0.05$). Thus, the extracted shape information differed significantly between women who did and did not develop GDM, and between subgroups of women diagnosed with GDM later in pregnancy, while the simple summary measures did not.

Article IV: Shape information in repeated glucose curves during pregnancy provided significant physiological information for neonatal outcomes.

The smoothed glucose curves at gestational weeks 14-16 and 30-32 showed large variation between the women at both visits. Glucose values were higher, and peaked later in third trimester than in early pregnancy.

In the multilevel FPCA, the first two subject-specific FPCs explained 98% of the variation across visits, and the first three subject- and visit-specific FPCs explained 92% of the residual variation within visits. Further analyses were restricted to these FPC curves and the corresponding FPC scores ($FPC1^{subj}$ and $FPC2^{subj}$, and $FPC1^{14\text{-}16}$, $FPC2^{14\text{-}16}$, $FPC3^{14\text{-}16}$, $FPC1^{30\text{-}32}$, $FPC2^{30\text{-}32}$ and $FPC3^{30\text{-}32}$, respectively). The dominating curve characteristic for the variation across visits ($FPC1^{subj}$) was "general glucose level", which accounted for 91% of this variation. The second most important curve characteristic across visits ($FPC2^{subj}$) was "timing of postprandial peak". The dominating curve characteristic for the residual variation within visits (accounting for 72% of this variation), was "general glucose level within visits", i.e. the general glucose level not accounted for by the general glucose level across visits. The second and third most important curve characteristics for the variation within visits were "timing of postprandial peak within visits" and "oscillating glucose within visits", respectively. These two characteristics accounted for a larger part of the variation within visits (15% and 8%), than across visits (7% and <2%).

There were strong correlations between $FPC1^{subj}$ scores and AUC at weeks 14-16 ($AUC^{14\text{-}16}$) and weeks 30-32 ($AUC^{30\text{-}32}$) (r=0.86 and 0.90, respectively), between $FPC1^{14\text{-}16}$ and $AUC^{14\text{-}16}$ (r=0.73), and between $FPC1^{30\text{-}32}$ and $AUC^{30\text{-}32}$ (r=0.87). All FPC1 scores were positively correlated with early pregnancy BMI ($0.12 \leq r \leq 0.35$).

Late postprandial peaks and/or high third trimester glucose levels had significant, positive effects on birth weight ($p<0.05$). Generally high glucose levels had a significant, positive impact on neonatal percentage fat ($p=0.04$). In addition, women with generally late glucose peaks gave birth to neonates with a somewhat higher percentage fat. High glucose level in third trimester had a significant, positive impact on cord blood C-peptide ($p=0.004$). In addition, neonates of women with oscillating glucose curves had somewhat lower C-peptide levels than those with one glucose peak during the OGTT.

5 Discussion

The quality of a study is often considered in terms of the precision, and the validity of the estimated effects (17). Whereas precision refers to random error, validity is often separated into internal validity, whether the effect estimates are biased due to the way the data is collected, analysed and interpreted, and external validity, whether the results from the study may apply or be generalized to populations or groups outside the study sample (1). Bias is defined as systematic deviations from the true effects (1), and is present when the association between exposure and outcome is not in its entirety the result of the causal effect of exposure on outcome (130). In the following, we discuss the internal validity of our study in terms of selection bias, information bias and confounding (sections 5.1-5.3), the statistical methods (section 5.4), the findings in the articles (section 5.5) and finally, the external validity (section 5.6).

5.1 Selection bias

Selection bias is present if the estimated association among those selected for the analysis differ from the association among those eligible (130). Bias caused by differential selection into the study sample are often referred to as selection bias (17), although it may in some instances be considered a bias due to unmeasured confounders that are not controlled for in the analysis. Selection bias and confounding may therefore be considered as partially overlapping concepts (17). It has been suggested to use DAGs to differentiate confounding, i.e. common causes of exposure and outcome, from a more specific definition of selection bias, i.e. bias resulting from conditioning on common effects (130). Volunteer bias and missing data bias, i.e. the bias that is present if the study is restricted to those who volunteered to participate, or the analysis is restricted to subjects with complete data, fall within the latter definition (130).

Approximately two thousand women registered for obstetric care at Rikshospitalet each year during the study period (114). Women who were invited to participate were chosen on basis of a Scandinavian name, and excluded if they had multiparous pregnancies or one of the diseases in the list of exclusion criteria (Figure 5.1). Due to logistics at the clinics involved in the study, not all the eligible women could be invited or included in the study. Approximately one third of the women accepted the invitation, and about five women were included every week. It is unlikely that this caused a selection bias, as the restrictions in invitations and inclusions were solely based on practical implementation of the study.

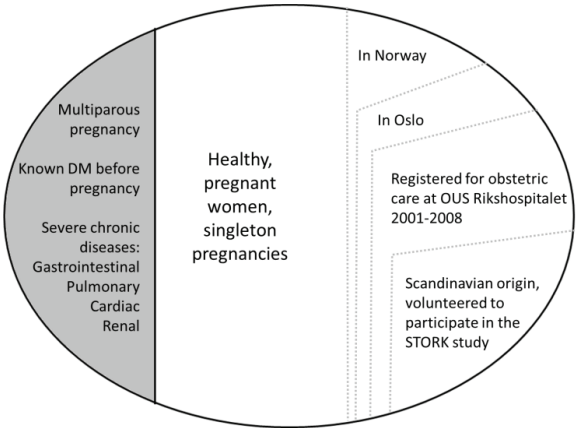


Figure 5.1 Overview of pregnant women and the exclusion and inclusion to the study. The grey area represents the exclusion criteria. The dotted grey lines show subgroups within healthy, pregnant women with singleton pregnancies.

Restricting the cohort to women of Scandinavian origin was done to avoid participants who could not understand the information that was given in the study, as this was only available in Norwegian. This reduced bias from unmeasured confounding due to genetic or epigenetic variation.

Volunteer bias may be present. For instance, a family history of DM may increase health awareness and affect the wish to participate in a study with close follow-up. Furthermore, the study's focus on big babies and birth complications may give a study sample with higher BMI values than in the eligible group. Then, as a family history of DM will possibly increase the risk of GDM, one may speculate whether the estimated association between BMI and GDM is slightly exaggerated in this thesis.

Furthermore, the recruitment to the cohort lasted for several years, and changes in the population characteristics during this period may have affected the volunteer bias unequally throughout the inclusion. For instance, repeated headlines in the media concerning the obesity epidemic, risk of type 2 DM or popular diets like the low-carb diet, may affect the pregnant population's behaviour and the motivation to participate in a study like the STORK study. However, it is not easy to evaluate the presence of such volunteer bias, or its impact on our estimates.

The analyses in articles I, III and IV were restricted to those with complete data. The sampling procedure to the sub-sample with inflammation data in article I included stratification and exclusion criteria based on birth weight, which was also the outcome in the

path analysis. The stratification is not likely to have caused selection bias. The exclusion was done to make sure that neonates with a deviating pathology should cause confounding, and can therefore be justified. In retrospect, it would have been better to use gestational age (less than 37 weeks) as the exclusion criterion, instead of birth weight. The FDA in article III was restricted to the 974 women (94%) with complete OGTT data from weeks 14-16, and the regression analysis with third trimester 2-h glucose as the outcome, to the 922 women (89%) who also had compete data on this and on early pregnancy BMI. Comparisons of the glucose values and BMI for women with incomplete OGTTs and/or missing BMI at weeks 14-16 and the study sample gave no statistically significant differences between the groups, although the mean values among the excluded were consequently slightly lower (not shown). It is therefore possible that the true values of the missing data also would have been in the lower range. The high proportion of complete data makes it is less likely that this would have an important impact on the results.

In article IV, the regression analysis of birth weight was restricted to the 868 women (84%) with term births and complete data on early pregnancy BMI, parity and OGTT data from weeks 14-16 and 30-32. Among those born at term, we found no statistically significant differences between those with complete data, and the registered data from those with incomplete data (results not shown), but these results may be obscured by the lack of information about the missing data. Again, the high proportion of complete data makes it is less likely the results for birth weight should be substantially biased. The regression analyses of neonatal percentage fat and cord blood C-peptide were done within participants with DXA data. C-peptide was not available from all neonates in this sub-sample, due to the amount of frozen cord blood. However, this was not related to the variables under study. Participants were recruited to DXA scanning during the study period, opening for an additional volunteer bias. The women in this sub-sample had a significantly lower BMI than the other women with term births in the STORK cohort (23.8 kg/m^2 vs 24.7 kg/m^2), and several significantly lower mean glucose values at weeks 14-16 and 30-32 (mean differences between the groups were in the range -0.1 mmol/l to 0.4 mmol/l). No significant difference in birth weight was found. The impact of these differences on the curve shape information extracted by FDA is difficult to assess. Also, the comparable birth weights make it less likely that this had an important impact on the results. This is supported by the results of the additional regression analyses of birth weight in the DXA sample, which gave similar results as in the large sample.

5.2 Information bias

Information bias occurs when the variables of interest, i.e. the main exposure, covariates and the outcome, are measured with measurement error. Measurement error in a categorical variable is often referred to as misclassification (121). Measurement error may be due to instrument error, i.e. the error related to measuring a specific quantity like blood glucose, at a specific point in time, and/or sampling error, i.e. errors occurring due to the process of obtaining the "true value" through sampling over time and/or space, like extracting glucose curve shape information from five discrete glucose measurements (121).

Measurements can have both random and systematic measurement error (1), and both may cause biased effect estimates (17;121). Random error may be heterogeneous, and has no apparent connection to another measurement or variable. Random error may be heterogeneous. Systematic error is error that is consistently wrong in a particular direction, and often has a recognisable source (1). Non-differential measurement error is error that does not depend on the outcome or other variables in the analysis, whereas differential measurement error is error that depends on the other variables in the analysis (17). Categorisation of a variable measured with random, non-differential error, will often give differential misclassification (131).

In the following, we will discuss measurement errors in the main exposures, mediators and outcomes. Measurement errors in covariates may also have contributed to information bias, but this is not discussed in detail.

De-trending of glucose values

We chose to use a point-of-care glucometer instead of a central laboratory to measure glucose, even though the glucometer was mainly recommended for screening purposes. There is always a trade-off between expenses and logistics, and the scientific accuracy and gold standard procedures. The glucometer had been used in various clinical settings at the hospital and several informal validations against the accredited laboratory had been found adequate (data not shown). However, there were no reports on the accuracy of the glucometer during long-term use. The unexpected trend in fasting glucose with a mean yearly increase of 0.11 mmol/l over seven years and the lack of a corresponding increase in the hexokinase data, the women's early pregnancy BMI, insulin or age, made it less likely that the observed trend in blood glucose values should have a biological cause, or be due to selection bias. Thus, glucose measurements were de-trended as described in article II.

Even though the error over time in fasting glucose measurements at inclusion was removed by the de-trending, there were still significant trends in the de-trended 2-h glucose values at inclusion, and in the fasting, 30 min and 2-h values at weeks 30-32 (Figure 1 in the Appendix). This may be a consequence of erroneous de-trending. However, the glucose measurements at inclusion were least likely to be affected by behavioural changes during pregnancy, and except from the 2-h values, all these measurements were successfully de-trended. We therefore trust the de-trending procedure to be adequate. The trends in the data from weeks 30-32 may represent a change in the population towards healthier glucose values during pregnancy, e.g. as a consequence of the increased focus on metabolic disorders, macrosomia and media focus on low-carb diets, or reflect selection bias, e.g. that an increasing proportion of the women included have a particularly healthy life style.

Categorisation of the 2-h glucose value in weeks 30-32

The measurement error in the de-trended 2-h glucose measurements at weeks 30-32, may cause differential misclassification in the categorised variable, which is difficult to evaluate. Misclassification due to random error, i.e. the instrumental error, often attenuate effect estimates (131).

Glucose curves

The individual curve fitting in articles III and IV was based on a measurement model with homogenous measurement error over the 2 hour time span, and smoothing with a roughness penalty. The degree of smoothing was based on an estimated, generalised cross-validation criterion, and the smoothing is assumed to reduce measurement error. Measurement error in glucose measurements may however be higher for high glucose values than for low, as indicated in the control chart in Figure 1 in the Appendix. The assumption of uniform measurement error may therefore be wrong. This issue was discussed during the analyses. However, as documentation and references for measurement errors are hard to find from the producers, and even harder to interpret in a clinical setting, we chose to keep the assumption. This is a non-differential measurement error. This may influence the smoothing of curves, and the smoothing procedure, e.g. more smoothing of higher values and less for lower values, may affect other results. The impact of the smoothing is discussed in the next paragraph.

FPC scores

The fitted curves were the basis for the FPCA, resulting in FPC curves and FPC scores. We found that FPCA of minimally smoothed curves gave FPC curves with a more wavy appearance, and a larger proportion of variance explained by FPC2 and FPC3 scores (results not shown). Hence, the smoothing of the curves did influence the information in the FPCs. In addition, in article IV, it is possible that the methods of covariance matrix estimation did not perfectly separate the across and within variances, and that the corresponding information extracted by the FPCA was slightly biased. Furthermore, leaving out the FPCs which explained the smallest part of the variation, i.e. those with the waviest appearance might have given a conservative estimate of the amount of curvature in the individual curves, which again could have caused bias in the FPC scores. This may affect the regression results, by giving a too high impact of the first FPCs ("general level") at the expense of the other FPCs ("timing of postprandial peak" and "oscillating curves"). It may also affect the variable selection in article IV by influencing the colinearity between the FPC scores.

BMI

Early pregnancy BMI was used as a surrogate for the women's general, non-pregnant body composition, as the latter was based on self-reported data. In weeks 14-16, the pregnancy may already have led to weight change: weight gain, or weight loss due to nausea and/or vomiting. It can also be discussed whether an assessment of the women's body composition in early pregnancy should have been done by the waist-hip ratio, but the same objections about changes in the women's constitution holds for these measurements. BMI was calculated from weight measured at inclusion and self-reported height. In retrospect, we regret that height was not measured, as self-reported height tend to be overestimated (132;133), leading to a systematic underestimation of BMI and differential misclassification of BMI categories. Differential measurement errors may also have occurred if those who gain less weight during the first weeks also overestimate their height most. However, we have little information in our data to investigate such possibilities, and it is difficult to predict the direction of the bias in the effect estimates.

Inflammatory markers

The validity of the inflammatory markers depends on whether we have measured the correct variables in the large and complex field of inflammation. If this were a psychometric study,

this issue would be part of the content validity, i.e. are the correct questions asked? Although a multi-disciplinary expertise participated in choosing inflammatory markers, the complexity of the topic makes it impossible to say with certainty that our study covered the inflammatory markers that represent the obesity-related markers of importance to pregnancy outcomes, or if the markers in article I are surrogate variables.

Measurements of inflammatory markers in blood samples are known to display large random variability, and infections at the time of measurement may affect the measurements. We therefore excluded women with a CRP above 10 mg/l. The dichotomisation of CRP may give differential misclassification. Women with high inflammatory response due to obesity and pregnancy may have been erroneously excluded due to infections, whereas women with lower inflammatory response may have been kept in the sample despite of infections that may have biased the measurements of both inflammation and glucose. The first error may lead to attenuated effect estimates. It is, however, difficult to evaluate the impact of measurement error on the estimates in the path model.

The neonatal outcomes birth weight, percentage fat and cord blood C-peptide

The measurement error of birth weight was reduced by using a calibrated weight, measuring the neonates shortly after birth.

DXA is being used increasingly as a reference method, and is considered the "gold standard" in body composition studies (93). It has, however, been demonstrated that error in DXA measurements is largest for the smallest subjects (134). It is less likely that this heterogeneous, non-differential measurement error will bias the results in our study, as all the neonates in the study sample were above 2315 g.

5.3 Confounding

Confounding is bias of the estimated effect of an exposure on an outcome due to the presence of a common cause of the exposure and the outcome (1). Confounding is an important issue in observational designs, and may lead to underestimation, overestimation, or even change the sign of the estimated effect (17). A confounder is a variable that is associated with the outcome (either as a cause or a proxy for a cause, but not as an effect of the disease), associated with the exposure, and not an effect of the exposure (17;121). The definition of confounding may also include bias due to baseline differences in exposure groups in the risk factor for the outcome, although this may be considered as selection bias (1), as in this thesis.

Confounding can be reduced by proper adjustment. Exploring data is not sufficient to identify whether a variable is a confounder, and such evaluation of confounding may lead to bias (1;15;17). Other evidence like pathophysiological and clinical knowledge and external data is needed. DAGs are useful tools when considering confounding variables (15;17).

Residual confounding

Confounding variables may be poorly measured or surrogate confounders, and others may be unmeasured. Also, the functional form of a regression model may be sub-optimal. The bias that remains after unsuccessful adjustment for confounders is called residual confounding (1;17). Residual confounding can never be ruled out in observational studies.

Mediators

In contrast to the confounder, a mediator represents a step in the causal pathway between the exposure and the outcome (1;17). Such a variable will also be associated with both the exposure and the outcome.

Confounding in articles I, III and IV

In article I, confounding and mediation were considered simultaneously. Even the simplified path model in Figure 3.3 included four regression equations (one for each of the outcome variables birth weight, fasting glucose, IL-1Ra and IL-6), and every arrow in the figure represents an effect that should be estimated without bias. Pregnancy complications may affect both the physiology of the mother during pregnancy, prematurity and birth weight of the child, and may thus be confounding variables. Our data included little information about pregnancy complications. We excluded those with a birth weight below the 10th percentile, thereby indirectly adjusting for several pregnancy complications. Women with infections (CRP>10 mg/l) at weeks 30-32 were also excluded, as altered levels of IL-6, IL-1Ra and fasting glucose in these women could be activated by other mechanisms than those related to BMI. This would therefore reduce bias from short-term infections at the time of the visit.

One can debate whether more variables representing common causes of the variables already in our model should have been added. For instance, life style factors, genetic factors, maternal age, parity, smoking, other cytokines, leptin, or changes in these values prior to weeks 30-32 may all potentially affect some of the variables in our model. Life style and genetic factors were unmeasured and cannot be adjusted for. It is, however, difficult to

evaluate the magnitude of this potential bias. Further, extending the model with variables that are measured requires causal knowledge, and implies specifications and biological justifications of all new hypothesized pathways. An extended model including more variables, both confounders and mediators, and with more detailed information about the hypothesized paths, may reduce residual confounding. As the literature is scarce concerning potential effects of variables like maternal age, smoking and parity on inflammatory markers during pregnancy, we chose to present a model without additional variables. We did additional adjustment for maternal age and smoking (only 5 women smoked) in the regression model for birth weight, but the results were similar (results not shown). Similar results were also found if gestational age was left out from the model (results not shown).

In article III, we estimated the effect of BMI category in early pregnancy on glucose curves in weeks 14-16 (by FANOVA), and the effect of glucose curves or simple OGTT summary measures in weeks 14-16 on the categorised 2-h glucose value in weeks 30-32 (by nominal logistic regression analysis). The FANOVA was performed without adjustment for covariates. As in article I, unmeasured life style and genetic factors may have influenced the results, but could not be adjusted for. It was decided to leave parity and maternal age out of the analyses for simplicity. Lack of adjustment for these variables or other potential confounders may have biased the results. Due to the degree of homogeneity of the study sample, the magnitude of such bias is likely to be small. Also, misclassification of BMI into categories (see section 5.2) may give bias due to residual confounding. We adjusted for early pregnancy BMI (continuous) in the nominal logistic regression analyses. Other covariates were not included, since it was beyond the scope of the article to build an extensive prediction model or to adjust for variables possibly on the causal pathway to the outcome. Curve shape information was incorporated in the multinomial logistic regression analysis by entering the independent FPC score variables in the analyses. Hence, as much as possible of the curve shape information was exploited in this analysis. The simple summary measures, on the other hand, contain less of the total information, and were included one at a time in the regression models, due to colinearity.

In article IV, early pregnancy BMI, parity and maternal age were included in the linear regression analysis of the effect of curve shape information on the neonatal outcomes birth weight, percentage fat and cord blood C-peptide. Weight gain may be on the causal pathway between the subject-specific glucose characteristics and birth weight, and was not included in the analyses. Again, unmeasured life style and genetic factors may have biased

the results, but the bias is likely of small magnitude due to the homogeneity of the study sample.

5.4 Discussion of statistical methods

The overall aim in this thesis was to use path analysis to analyse mediation, and FDA to analyse OGTT data, to better exploit important physiological information in the STORK data. Below we discuss possible improvements and alternative approaches.

Path analysis (article I)

The hypothesized path diagram, in which early pregnancy BMI leads to increased inflammation with secondary effects on glucose regulation and fetal growth, is a simplified model of the complete process. Integrative physiology is much more complex, and studies of larger, and other, models should be performed. For instance, TNF α might play a role in insulin resistance in pregnancy (47). However, results are partly conflicting concerning its association with BMI (135), and TNF α was not included in our analysis. Further, MCP-1 may also play a role in metabolism. A biological justification of pathways for MCP-1 in a path model together with IL-6 and IL-1Ra was not found in the literature, and a simplified model without MCP-1 was implemented. While there are certain pitfalls that can cause flawed conclusions in a path analysis, this method can be justified in problems with a simple structure and linear relations between variables (2). We used standardized variables and focused on comparing the relative importance of variables in our data. Using the original variables instead may ease interpretation when comparing our results with other studies.

The Bayesian model

Effect estimates in this study could have been obtained with a frequentistic analysis, which is the dominant analytical approach in clinical research, but we chose a Bayesian approach. In a frequentistic analysis, parameters are assumed to be fixed, but unknown quantities, and must be estimated from the data. Potential prior knowledge is treated informally in the interpretation of the results. In Bayesian analysis, in contrast, all parameters are assumed to be stochastic variables, and effect estimates and corresponding CrIs are derived from estimated probability distributions. It is therefore necessary to specify the a priori knowledge of the situation and the parameters in terms of a statistical distribution of the parameters under study.

Frequentistic analyses are based on normality assumptions and central limit theory, whereas Bayesian analyses rely on prior assumptions and simulation techniques. Bayesian analysis is computation heavy, and technically more complicated, but the WinBUGS software has made Bayesian methods available, also to clinical researchers (128;136). Bayesian models have the strength that they can handle non-normality and non-linearity more easily than traditional analyses, and they are also flexible with respect to several types of variables (128;136). In studies of complex biological mechanisms, where samples will typically be small due to the costs and restraints in data collecting, data tend to be skewed and where there are non-linear relations (137-139), Bayesian methods represent a valuable tool

De-trending (article II)

The glucose data from the observational period was de-trended based on a weighted average of the results from two linear regression analyses, i.e. assuming a globally applicable model.

Figure 1 in the article (left, upper plot, as well as Figure 1 in the Appendix) indicate changes in the hyperglycaemic controls after approximately 2.5, 4, and 6 years (between observations 101 and 151 in Figure 1 in the Appendix) that may be interpreted as shifts, not trends (122). Such shifts were not equally apparent in the hypoglycaemic control. An important question was whether these shifts were actually due to reagent strip lot changes or control lot changes, i.e. alterations of the measurement process caused by the study group. The study records showed that mainly four different lots of glucose strips were used during the observational period, and these were changed approximately 1.2, 3.0, and 6.0 years after study start. In the period 1.1-1.3 years after study start, strips from three different lots were used. That is, except from the change at 6.0 years, lots of glucose strips changed at different times than the apparent shifts in the data in Figure 1. Information on changes in control lots was unfortunately only available for the last part of the study period, and not from the period where a shift was most pronounced, i.e. after 2.5 years.

In summary, the available dates for change in strips or controls could not explain the observed shifts or trends. Shifts were not identified consecutively during the study period, and underlying reason(s) for shifts could not be identified retrospectively. We therefore chose to de-trend the data by an overall procedure.

Categorisation of continuous variables (article III)

In order to ease the presentation of FDA for a non-technical audience, we categorised early pregnancy BMI for the FANOVA, and used the categorised 2-h OGTT value in weeks 30-32 as the outcome in nominal logistic regression. From a statistical point of view, such categorization is not usually recommended (140;141), and functional regression with BMI as a continuous explanatory variable, and linear or non-linear regression analysis with the 2-h value as a continuous outcome, would be preferable.

FDA (articles III and IV)

In the literature, visualisations of "mean glucose curves" are usually presented as means at selected time points with interpolation lines, and variability is usually quantified by SDs or standard errors at the same time points (142-144). While previous studies of shape information from glucose curves have focused on either simple shape indices or advanced parametric modelling (65-68), we have used statistical tools developed specifically for analysing curve data.

The scarce sampling of glucose during the OGTT is likely to obscure the extraction and interpretation of the curve characteristics. More physiologically interesting temporal details and better discriminating abilities of the FPCs may be expected in a more heterogeneous population than in our study sample, and from OGTT curves over more than 2 hours or with a more frequent OGTT sampling (68). With more measurements per OGTT, it is also possible to apply alternative smoothing strategies (125;126).

Consistent with results from other studies, AUC was much better than the widely used fasting glucose or 2-h value in capturing the essential temporal information of OGTT glucose curves (145-147).

In article III, the strongest association between the shape index defined in (66) and the FPC scores was found for FPC3 scores which explained the smallest proportion of the total variance. These scores should be interpreted with caution in our study, but might explain a larger part of the total variation in studies with more frequent sampling. The continuous FPC3 score variable provided quantification of curvature, which is preferable in order to retain both temporal information and statistical power (140). The shape index defined in (66), in contrast, is based on an a priori classification of curves into the categories "biphasic", "monophasic" or "unclassified", and involves several ad hoc thresholds for the categorisation (66). Many curves (27%) failed to meet the classification criteria and were left out of the

analyses, resulting in a severe reduction of power and a biased representation of metabolic profiles in the study sample. Another, recently suggested shape index (67) is based on a rough approximation of the mean of the second order derivatives in the intervals between the measurements during the OGTT, giving a rough approximation of the total curvature. Our results showed that curvature is better extracted by FDA.

Parametric modelling based on differential equation models of physiological mechanisms is an alternative approach to the analysis of full glucose curves (68;148-150). A major disadvantage of parametric models is that estimating each person's individual parameters requires many measurements, often based on intravenous test procedures (151). The OGTT is the simplest and most frequently used test procedure in larger studies because intravenous procedures such as the euglycaemic clamp (58) are time-consuming, invasive and labour intensive. The data-driven approach of FDA is well-suited for the analysis of glucose regulation in larger studies.

Longitudinal data analysis with five repeated measurements per OGTT, and random effect of woman and modelling of the covariance structure is an alternative approach. It is also possible to use ordinary PCA scores based on the five glucose variables from weeks 14-16 as input to the regression analysis of glucose tolerance later in pregnancy, instead of scores from FPCA. With only five measurements per curve and measurements taken at the same time points for each woman, such traditional multivariate methods would be expected to extract similar information as the FDA. However, all these methods are similar in the sense that they approach the curve data only indirectly, by applying techniques originally developed for other types of data. FDA has its strength in being developed for analysing such data directly, and, in addition to being a more intuitively applicable methodology, it emphasizes the basic assumption about continuity of the underlying process, is easier to apply in situations with more frequent sampling and sampling at unequal time points.

The impact of the smoothing procedure on the results was mentioned in section 5.2. The multilevel models described in (125;126) explain how FPCA of smoothed covariance surfaces can be used as the first step in the analysis. This means that the B-spline smoothing described in (24), is replaced with smoothing of covariance surfaces (126). In our data, due to the few measurements, such a smoothing procedure failed to give realistic estimates (results not shown), and we therefore chose individual curve B-spline smoothing in article IV. For first-time users of FDA, the approach with smoothed covariance surfaces is not as

immediately intuitive as the curve-fit approach, and we therefore consider the latter to be an advantage in the communication with clinical researchers.

5.5 Discussion of findings

Inflammatory markers as mediating variables in the relation between early pregnancy BMI and birth weight (article I)

The hypothesized associations were partly confirmed in our data. According to the results from our simplified model, about 20% of the effect of early pregnancy BMI on birth weight was mediated through paths involving IL-1Ra. However, IL-1Ra is a dual marker; it is an anti-inflammatory cytokine, binding to IL-1 receptor without inducing an effect, but at the same time reflects an activation of the IL-1 system and is also a marker of inflammation in general (49;152). Based on this, we cannot rule out that the measured effect of IL-1Ra reflects the action of IL-1β. Although results from observational studies must be interpreted with caution concerning causality, our results indicate a substantial role for the interleukin 1-system in the deranged glucose metabolism associated with high maternal BMI during pregnancy and consequently an important role for interleukins as mediators between maternal fat-mass, glucose and birth weight.

The use of birth weight as a marker of fetal growth might explain why our results were not in accordance with the previously reported association between maternal IL-6 and prenatal growth reflected by neonatal fatmass (153).

Despite the large sample size, we did not find significant direct effects of the interleukins on birth weight, possibly because IL-6 and IL-1Ra do not play an important role in regulating fetal growth through changing placental properties. The result may also be due to the fact that cytokines display pleiotropic effects and show considerable biological variation (154). Furthermore, moderate effect estimates were anticipated. As basis for comparison, maternal BMI, one of the major determinants of birth weight, accounts for less than 15% of the variation in birth weight in other studies (13;155;156). Our BMI result was similar, and borderline significant, possibly due to the homogeneity of our study sample and a lack of power to detect small effects. Based on these considerations, we reported our model with all the original arrows present, and indirect effects were estimated with significant and non-significant direct effects included although not all the direct effects were significant.

We chose two markers as representatives of the inflammatory status in obese women, being aware that other markers may be important as mediators in the association between

BMI and birth weight. In addition, effects of cytokines on birth weight are probably not an effect of a single mediator, but rather the result of the interactions of several and in combinations (154;157). Therefore, we cannot rule out that cytokines in combinations may have a direct effect on placental properties and birth weight even if we were not able to find such an effect.

Shape information inherent in glucose curves in pregnancy (articles III and IV)

The mean glucose curve obtained from FDA corresponded well with the familiar shape of glucose curves (142-144). Individual glucose curves have been presented in several publications (68;144;158), but the variability in curve trajectories is highly under-reported, and thus largely unknown. The information indicated by the shape of glucose curves is therefore rarely used in clinical practice, and only occasionally in research. Our findings emphasise the large variability in glucose curves, even though the study sample was relatively homogenous, and provide a reference for glucose curves in healthy, pregnant women.

The interpretation of FPC curves is essential for the usefulness of FPCA. Current insight into metabolism supported the interpretations of the FPCs ("general glucose level", "timing of postprandial peak" and "oscillations") as plausible and important physiological features. The identification of the general glucose levels as the most dominant characteristics of individual glucose curves was supported by the strong associations between FPC1 scores and the AUCs, and the positive association with early pregnancy BMI. The association with BMI is in accordance with physiological knowledge of obesity and insulin resistance (51;159). The importance of the general glucose level is also in accordance with numerous studies focusing on the importance of elevated glucose of various types, e.g. fasting, 1-h, 2-h or HbA1c values, in diabetes research (51;69;79).

In article IV, we found that the general glucose level accounted for a larger part of the variation across, than within visits, whereas the timing of the peak was more important for the variation within, than across visits. This is not surprising, as FPC2 represent more curvature than FPC1, and some of the curvature may be averaged out at the subject-specific level.

The elevated postprandial levels in third trimester, the small increase in fasting glucose, and the large increase in fasting insulin and prevalence of GDM from inclusion to weeks 30-32, are in accordance with an expected progressive insulin resistance among

pregnant women, and supports the finding of timing of postprandial peak as the second most important curve characteristic (51). The increase in postprandial values during pregnancy, and corresponding delay in postprandial peak, is also supported by several earlier studies (60-62;64;160). Many studies have found a decline in fasting glucose during the first trimester of pregnancy (53), but an overview of longitudinal studies during pregnancy showed conflicting results concerning later pregnancy fasting glucose (53). This justifies our findings of a small increase in fasting glucose from weeks 14-16, to 30-32.

The interpretation of FPC3s as "oscillations" was chosen on basis of physiological theories and studies with more frequent sampling during OGTTs (68;161).

In article III, the FPC1 scores, 2-h values and AUC at weeks 14-16 differed significantly between groups of women without a GDM diagnosis at weeks 30-32. However, only FPC2 scores at weeks 14-16 were significantly different between women with and without GDM at weeks 30-32 and only FPC2 and FPC3 scores differed significantly between diabetic women with the highest and second highest 2-h values in the third trimester. Thus, at weeks 14-16, FPC1 or AUC alone did not capture all of the essential information about the differences in glucose metabolism. To distinguish curve trajectories reflecting deviating glucose tolerance from those considered normal, the information from FPC2 and FPC3 was necessary. A study of type 1 DM patients with islet transplantations found that increased glucose AUC and time to peak C-peptide after metabolic testing were metabolic markers of islet allograft dysfunction (162), supporting the physiological importance of both FPC1 and FPC2 scores. The timing of the peak C-peptide was also found to be predictive of progression to type 1 DM in the Diabetes Prevention Trial (163).

The HAPO study findings include significantly higher odds ratios for high birth weight, cord-blood serum C-peptide level and percentage body fat (above their respective 90th percentiles), for high fasting, 1-hour and 2-hour glucose levels (79;81), which supports our findings of important impact of FPC1 scores, interpreted as "general glucose level", on these outcomes. Other studies with a similar scope, but smaller sample sizes are also in accordance with these findings (164-168). However, none of these studies addressed the impact of the dynamic regulation of the blood glucose, which is embedded in the FPC2 and FPC3 scores. Some studies have commented on the postprandial peak and birth outcomes (61;169;170), but to our knowledge, our study is the first to formally investigate the impact of the timing of the postprandial peak. The extension of the Pedersen hypothesis to the normal-glycaemic range (79), and the new GDM criteria which takes into account both the fasting, 1-

h and 2-h OGTT values (171) indirectly also supports the findings of an important role of dynamics in the curves.

5.6 External validity

Inclusion and exclusion criteria make the population of eligible subjects relatively homogenous and make it easier to argue for good internal validity, as the lack of distinct sub-groups, e.g. due to genetic or epigenetic variation, is likely to reduce the number of confounders. On the other hand, results from a strongly selected sample may be difficult to generalise to subjects outside the study sample. Restricting the cohort to women of Scandinavian origin (Figure 5.1) makes the study sample less representative for the total population in Norway, and especially in Oslo, where there is a relatively large proportion (approximately 20%) of non-Scandinavian immigrants (172).

The overall aim of the STORK study was to gain insights into maternal metabolic syndrome and determinants of foetal macrosomia (114). It can be argued that the general physiological mechanisms studied in our articles will be similar in all healthy pregnant women. However, variation between populations has been found, i.e. for the associations between pre-pregnancy BMI and the risk of GDM (46), and for the associations between maternal obesity and pregnancy outcomes (39). Also, glucose values and the prevalence of DM and GDM vary substantially between populations (74;75). This may be a consequence of genetic, social or life style factors. Such factors are not measured in our study, but due to the selection procedure, they may be expected to vary less in our study samples than among pregnant women in general. Hence, it cannot be ruled out that some of our results will be different in other populations. Future studies are needed to validate our findings and to evaluate the extent of this difference, and whether associations are weaker or stronger in other populations.

Participation in the study might lead to behavioural change during pregnancy, and make the results from late pregnancy less generalizable to pregnant women outside the study. A comparable study of pregnant women found that there was a significant reduction in smoking, alcohol consumption and intake of caffeinated drinks, but little change in fruit and vegetable intake (173). Thus, behavioural changes are not likely to be substantial, and the data from weeks 30-32 are as generalizable as the early pregnancy data.

6 Conclusions

We have used path analysis and FDA to gain physiological insight into problems of clinical interest in a cohort of pregnant women. Path analysis is particularly suited for the analysis of mediation. Although it is far from being a modern statistical method, it is under-utilised in some areas of clinical research, and our contribution is therefore important. FDA and multilevel functional models represent novel approach to the analysis of glucose measurements, and for analysing functional data at several visits in relation to an outcome. We have facilitated the presentation of the analyses and results, to make such analyses available to applied statisticians and clinical researchers.

More specifically, we found that

- The use of path analysis in combination with current biological concepts added to the knowledge of the adipose tissue-derived inflammatory factors IL-6 and IL-1Ra as mediators in the association between early pregnancy BMI and birth weight. Even though the metabolic pathways are complex, simplified path models like the one in article I may be useful.
- The unexpected trend during long-term use of a glucose measurement system most likely originated from the measurement system. We successfully removed the trend by a weighted average of regression estimates based on independent control solutions.
- FDA of glucose curves in early pregnancy was superior to traditional analyses of OGTT data in terms of providing physiologically interpretable and important temporal information, specifically in terms of differentiating between women who did and did not develop GDM later in pregnancy.
- The shape information from glucose curve data from two time points during pregnancy had significant impact on birth weight, neonatal percentage of fat, and C-peptide in cord blood, demonstrating physiological relevance of a late postprandial glucose peak, generally high glucose levels during pregnancy, and high third trimester glucose levels. Shape information inherent in entire glucose curves is important for several outcomes, and may contribute to the understanding of the metabolic changes during pregnancy.

7 Clinical implications and future perspectives

The topics of observational clinical studies often include problems with complex structures, and statistical methods other than those traditionally applied in clinical research may be necessary to answer the research questions. For instance, path analyses, and recently also FDA, are methods that are well known from other research fields, but such analyses are still not common knowledge or common tools in clinical studies. As an example, one of the reviewers of article I commented the lack of *p* values. To statisticians, Bayesian inference based on estimates and CrIs is a matter of convenience, but clinical readers may find the interpretation of the results challenging. Applied biostatisticians should take on this task.

The results from article I imply that there is a need for experimental studies to assess the molecular mechanisms of inflammation in relation to obesity and pregnancy.

We found that a point-of-care glucose device may give erroneous results when used over a long time period. Our results should be taken into consideration when such devices are developed, produced and used. The results in article II emphasise the importance of quality control of measurement procedures. Health care workers and researchers must be aware of the need to perform regular quality controls, including the use of control charts and regularly sampled blood tests, analysed at an accredited laboratory (56;122). Such procedures will unmask trends or shifts during the study period and make it possible to search for underlying reasons, like changes in reagent strip lots or control lots.

We recommend the FDA approach for the analysis of glucose data sampled repeatedly during glucose tolerance testing, or continuous glucose monitoring, to capitalize on important information that would otherwise be lost. Continuous glucose monitoring might increasingly be used in future studies and in individual patient care to obtain OGTT measurements and measurements of glucose profiles in daily life. Currently, many studies of data from continuous glucose monitoring devices restrict the analyses to simple summary measures like the mean glucose (63;174). Such data should be analysed by FDA. Challenges in the analysis of data from daily life glucose profiles will include curve alignment (26), which was not a topic in our analyses.

Furthermore, comparison of curve shape information from individuals with insulin resistance or beta cell failure might reveal whether curve features can distinguish between these two main processes that lead to the development of diabetes.

The AUC was strongly correlated with FPC scores that provided information about the general glucose level during the OGTT, but not with scores providing information about

timing of postprandial peak or oscillations, nor with the fasting or 2-h values. Our recommendation is therefore to use AUC values rather than the fasting values or 2-h values, if FDA is not applied.

The presented techniques should also be explored in studies in non-pregnant populations. Examples include studies of metabolic disorders, metabolic changes during or after meals or after physical exercise, and diurnal measurements of hormone regulation. This is in accordance with a recent review (32).

Shape information inherent in glucose curves can contribute to a better understanding of the different stages in the development of unhealthy glucose metabolism, and to a more precise prediction of women at risk for maternal or foetal complications. Then, interventions targeted to modify glucose curves could be initiated before a GDM diagnosis is given, or treatment for it is necessary. Such interventions have been studied in pregnant and non-pregnant study samples (175-177). Future studies should investigate whether such interventions also may affect pregnancy outcomes, and have positive long-term effects on maternal health.

References

(1) Porta M. A dictionary of epidemiology. 5th ed. New York: Oxford University Press; 2008.

(2) Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. Int J Epidemiol 2013;42:1511-9.

(3) Stark MJ, Clifton VL, Wright IMR. Carbon Monoxide is a Significant Mediator of Cardiovascular Status Following Preterm Birth. Pediatrics 2009;124:277-84.

(4) Jansson N, Nilsfelt A, Gellerstedt M, Wennergren M, Rossander-Hulthen L, Powell TL, et al. Maternal hormones linking maternal body mass index and dietary intake to birth weight. Am J Clin Nutr 2008;87:1743-9.

(5) Pilgaard K, Faerch K, Carstensen B, Poulsen P, Pisinger C, Pedersen O, et al. Low birthweight and premature birth are both associated with type 2 diabetes in a random sample of middle-aged Danes. Diabetologia 2010;53:2526-30.

(6) Gangestad SW, Caldwell Hooper AE, Eaton MA. On the function of placental corticotropin-releasing hormone: a role in maternal-fetal conflicts over blood glucose concentrations. Biol Rev Camb Philos Soc 2012;87:856-73.

(7) Godfrey KM. The role of the placenta in fetal programming-a review. Placenta 2002;23 Suppl A:S20-S27.

(8) King JC. Maternal obesity, metabolism, and pregnancy outcomes. Annu Rev Nutr 2006;26:271-91.

(9) Armitage P, Colton T. Encyclopedia of Biostatistics. In *Path analysis* (ed. Bollen KA), Vol. 4, pp. 3280-3284. John Wiley & Sons; 1998.

(10) Ditlevsen S, Christensen U, Lynch J, Damsgaard MT, Keiding N. The mediation proportion: a structural equation approach for estimating the proportion of exposure effect on outcome explained by an intermediate variable. Epidemiology 2005;16:114-20.

(11) Bodnar LM, Ness RB, Harger GF, Roberts JM. Inflammation and triglycerides partially mediate the effect of prepregnancy body mass index on the risk of preeclampsia. Am J Epidemiol 2005;162:1198-206.

(12) Gamborg M, Andersen PK, Baker JL, Budtz-Jorgensen E, Jorgensen T, Jensen G, et al. Life Course Path Analysis of Birth Weight, Childhood Growth, and Adult Systolic Blood Pressure. Am J Epidemiol 2009;169:1167-78.

(13) Fleten C, Stigum H, Magnus P, Nystad W. Exercise during pregnancy, maternal prepregnancy body mass index, and birth weight. Obstet Gynecol 2010;115:331-7.

(14) Kruger DJ, Clark J, Vanas S. Male Scarcity is Associated with Higher Prevalence of Premature Gestation and Low Birth Weight Births Across the United States. Am J Hum Biol. 2013;25:225-7.

(15) Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology. Am J Epidemiol 2002;155:176-84.

(16) Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10:37-48.

(17) Rothman KJ, Greenland S, Lash TL. Modern Epidemiology, 3rd ed. Lippincott Williams & Wilkins; 2008.

(18) Sutherland JP, McKinley B, Eckel RH. The metabolic syndrome and inflammation. Metab Syndr Relat Disord 2004;2:82-104.

(19) Bertram R, Sherman A, Satin LS. Metabolic and electrical oscillations: partners in controlling pulsatile insulin secretion. Am J Physiol Endocrinol Metab 2007;293:E890-E900.

(20) Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis Model Assessment - Insulin Resistance and Beta-Cell Function from Fasting Plasma-Glucose and Insulin Concentrations in Man. Diabetologia 1985;28:412-9.

(21) Kostovski E, Iversen PO, Birkeland K, Torjesen PA, Hjeltnes N. Decreased levels of testosterone and gonadotrophins in men with long-standing tetraplegia. Spinal Cord 2008;46:559-64.

(22) Lovstad M, Funderud I, Lindgren M, Endestad T, Due-Tonnessen P, Meling T, et al. Contribution of subregions of human frontal cortex to novelty processing. J Cogn Neurosci 2012;24:378-95.

(23) Ramsay JO, Hooker G, Gray J. Functional data analysis with R and MATLAB. Springer; 2009.

(24) Ramsay JO, Silverman BW. Functional data analysis. 2nd edition. Springer; 2005.

(25) Ramsay JO. Functional data analysis. Available from http://www.functionaldata.org

(26) Sorensen H, Goldsmith J, Sangalli LM. An introduction with medical applications to functional data analysis. Stat Med 2013;32:5222-40.

(27) Trail JB, Collins LM, Rivera DE, Li R, Piper ME, Baker TB. Functional Data Analysis for Dynamical System Identification of Behavioral Processes. [Epub ahead of print] Psychol Methods 2013

(28) Viviani R, Gron G, Spitzer M. Functional principal component analysis of fMRI data. Human Brain Mapping 2005;24:109-29.

(29) West RM, Harris K, Gilthorpe MS, Tolman C, Will EJ. Functional data analysis applied to a randomized controlled clinical trial in hemodialysis patients describes the variability of patient responses in the control of renal anemia. Journal of the American Society of Nephrology 2007;18:2371-6.

(30) Coffey N, Harrison AJ, Donoghue OA, Hayes K. Common functional principal components analysis: A new approach to analyzing human movement data. Hum Mov Sci 2011;30:1144-66.

(31) Duhamel A, Devos P, Bourriez JL, Preda C, Defebvre L, Beuscart R. Functional data analysis for gait curves study in Parkinson's disease. Stud Health Technol Inform 2006;124:569-74.

(32) Ullah S, Finch CF. Applications of functional data analysis: A systematic review. BMC Med Res Methodol 2013;13:43.

(33) World Health Organization. Obesity: Preventing and managing the global epidemic. Report of a WHO Consultation. WHO Technical Report Series 894. World Health Organization; 2000.

(34) Sewell MF, Huston-Presley L, Amini SB, Catalano PM. Body mass index - A true indicator of body fat in obese gravidas. J Reprod Med 2007;52:907-11.

(35) Lindsay CA, Huston L, Amini SB, Catalano PM. Longitudinal changes in the relationship between body mass index and percent body fat in pregnancy. Obstet Gynecol 1997;89:377-82.

(36) World Health Organization. Obesity and overweight. Fact sheet N°311, Updated March 2013 Available from http://www.who.int/mediacentre/factsheets/fs311/en/index.html

(37) Hull HR, Dinger MK, Knehans AW, Thompson DM, Fields DA. Impact of maternal body mass index on neonate birthweight and body composition. Am J Obstet Gynecol 2008;198:416.

(38) Catalano PM, Ehrenberg HM. The short- and long-term implications of maternal obesity on the mother and her offspring. BJOG 2006;113:1126-33.

(39) Nelson SM, Matthews P, Poston L. Maternal metabolism and obesity: modifiable determinants of pregnancy outcome. Hum Reprod Update 2010;16:255-75.

(40) World Health Organization. Global health risks. Mortality and burden of disease attributable to selected major risks. World Health Organization; 2009.

(41) Metzger BE. Hyperglycaemia and Adverse Pregnancy Outcome (HAPO) Study: associations with maternal body mass index. BJOG 2010;117:575-84.

(42) Callaway LK, Prins JB, Chang AM, McIntyre HD. The prevalence and impact of overweight and obesity in an Australian obstetric population. Med J Aust 2006;184:56-9.

(43) Cnattingius S, Villamor E, Lagerros YT, Wikstrom AK, Granath F. High birth weight and obesity--a vicious circle across generations. Int J Obes (Lond) 2012;36:1320-4.

(44) Reynolds RM, Allan KM, Raja EA, Bhattacharya S, McNeill G, Hannaford PC, et al. Maternal obesity during pregnancy and premature mortality from cardiovascular event in adult offspring: follow-up of 1 323 275 person years. BMJ 2013;347:f4539.

(45) Cai D, Yuan M, Frantz DF, Melendez PA, Hansen L, Lee J, et al. Local and systemic insulin resistance resulting from hepatic activation of IKK-beta and NF-kappaB. Nat Med 2005;11:183-90.

(46) Torloni MR, Betran AP, Horta BL, Nakamura MU, Atallah AN, Moron AF, et al. Prepregnancy BMI and the risk of gestational diabetes: a systematic review of the literature with meta-analysis. Obes Rev 2009;10:194-203.

(47) Richardson AC, Carpenter MW. Inflammatory mediators in gestational diabetes mellitus. Obstet Gynecol Clin North Am 2007;34:213-24.

(48) Lash GE, Ansari T, Bischof P, Burton GJ, Chamley L, Crocker I, et al. IFPA Meeting 2008 Workshops Report. Placenta 2009;30:S4-S14.

(49) Juge-Aubry CE, Henrichot E, Meier CA. Adipose tissue: a regulator of inflammation. Best Pract Res Clin Endocrinol Metab 2005;19:547-66.

(50) Ramsay JE, Ferrell WR, Crawford L, Wallace AM, Greer IA, Sattar N. Maternal obesity is associated with dysregulation of metabolic, vascular, and inflammatory pathways. J Clin Endocrinol Metab 2002;87:4231-7.

(51) Hod M, Jovanovic L, Di Renzo GC, de Leiva A, Langer O. Textbook of diabetes and pregnancy. Taylor & Francis; 2008.

(52) Lain KY, Catalano PM. Metabolic changes in pregnancy. Clinical Obstetrics and Gynecology 2007;50:938-48.

(53) Mills JL, Jovanovic L, Knopp R, Aarons J, Conley M, Park E, et al. Physiological reduction in fasting plasma glucose concentration in the first trimester of normal pregnancy: The diabetes in early pregnancy study. Metabolism. 1998;47:1140-4.

(54) Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 1997;20:1183-97.

(55) Buhling KJ, Henrich W, Kjos SL, Siebert G, Starr E, Dreweck C, et al. Comparison of point-of-care-testing glucose meters with standard laboratory measurement of the 50g-glucose-challenge test (GCT) during pregnancy. Clin Biochem. 2003;36:333-7.

(56) Cull CA, Manley SE, Stratton IM, Neil HAW, Ross IS, Holman RR, et al. Approach to maintaining comparability of biochemical data during long-term clinical trials. Clin Chem 1997;43:1913-8.

(57) Michigan Diabetes Research and Training Center. Hemoglobin A1c Fact Sheet. 2013. Available from http://www.med.umich.edu/mdrtc/cores/ChemCore/hemoa1c.htm

(58) Defronzo RA, Tobin JD, Andres R. Glucose Clamp Technique - Method for Quantifying Insulin-Secretion and Resistance. Am J Physiol 1979;237:E214-23.

(59) Davidson MB. Counterpoint: The oral glucose tolerance test is superfluous. Diabetes Care 2002;25:1883-5.

(60) Catalano PM, Tyzbir ED, Wolfe RR, Calles J, Roman NM, Amini SB, et al. Carbohydrate-Metabolism During Pregnancy in Control Subjects and Women with Gestational Diabetes. Am J Physiol 1993;264:E60-E67.

(61) Parretti E, Mecacci F, Papini M, Cioni R, Carignani L, Mignosa M, et al. Third-trimester maternal glucose levels from diurnal profiles in nondiabetic pregnancies: correlation with sonographic parameters of fetal growth. Diabetes Care 2001;24:1319-23.

(62) Lind T, Billewic WZ, Brown G. Serial Study of Changes Occurring in Oral Glucose-Tolerance Test During Pregnancy. J Obstet Gynaecol Br Commonw. 1973;80:1033-9.

(63) Siegmund T, Rad NT, Ritterath C, Siebert G, Henrich W, Buhling KJ. Longitudinal changes in the continuous glucose profile measured by the CGMS in healthy pregnant women and determination of cut-off values. Eur J Obstet Gynecol Reprod Biol 2008;139:46-52.

(64) Hadden DR, McLaughlin C. Normal and abnormal maternal metabolism during pregnancy. Semin Fetal Neonatal Med 2009;14:66-71.

(65) Zhou WB, Gu YY, Li H, Luo M. Assessing 1-h plasma glucose and shape of the glucose curve during oral glucose tolerance test. Eur J Endocrinol 2006;155:191-7.

(66) Tschritter O, Fritsche A, Shirkavand F, Machicao F, Haring H, Stumvoll M. Assessing the shape of the glucose curve during an oral glucose tolerance test. Diabetes Care 2003;26:1026-33. Erratum 27:1855.

(67) Tura A, Morbiducci U, Sbrignadello S, Winhofer Y, Pacini G, Kautzky-Willer A. Shape of glucose, insulin, C-peptide curves during a 3-h oral glucose tolerance test: any relationship with the degree of glucose tolerance? Am J Physiol Regul Integr Comp Physiol 2011;300:R941-R948.

(68) Trujillo-Arriaga HM, Roman-Ramos R. Fitting and evaluating the glucose curve during a quasi continuous sampled oral glucose tolerance test. Comput Biol Med 2008;38:185-95.

(69) Alberti KGMM, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications part 1: Diagnosis and classification of diabetes mellitus - Provisional report of a WHO consultation. Diabet Med 1998;15:539-53.

(70) O'Sullivan JB, Mahan CM. Criteria for the oral glucose tolerance test in pregnancy. Diabetes 1964;13:278-85.

(71) O'Sullivan JB. Establishing criteria for gestational diabetes. Diabetes Care 1980;3:437-9.

(72) Knopp RH. John B. O'Sullivan: a pioneer in the study of gestational diabetes. Diabetes Care 2002;25:943-4.

(73) Carpenter MW, Coustan DR. Criteria for screening tests for gestational diabetes. Am J Obstet Gynecol 1982;144:768-73.

(74) World Health Organization. Diagnostic Criteria and Classification of Hyperglycaemia First Detected in Pregnancy. World Health Organization; 2013.

(75) Wendland EM, Torloni MR, Falavigna M, Trujillo J, Dode MA, Campos MA, et al. Gestational diabetes and pregnancy outcomes--a systematic review of the World Health Organization (WHO) and the International Association of Diabetes in Pregnancy Study Groups (IADPSG) diagnostic criteria. BMC Pregnancy Childbirth 2012;12:23.

(76) Pedersen J. Weight and length at birth of infants of diabetic mothers. Acta Endocrinol (Copenh) 1954;16:330-42.

(77) Lindsay RS. Many HAPO returns: maternal glycemia and neonatal adiposity: new insights from the Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) study. Diabetes 2009;58:302-3.

(78) Contreras M, Sacks DA, Watson W, Dooley SL, Foderaro M, Niznik C, et al. The Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study. Int J Gynaecol Obstet. 2002;78:69-77.

(79) Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, et al. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008;358:1991-2002.

(80) Coustan DR, Lowe LP, Metzger BE, Dyer AR. The Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study: paving the way for new diagnostic criteria for gestational diabetes mellitus. Am J Obstet Gynecol 2010;202:654-6.

(81) Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study: associations with neonatal anthropometrics. Diabetes 2009;58:453-9.

(82) Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, Damm P, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010;33:676-82.

(83) American Diabetes Association. Standards of Medical Care in Diabetes-2011. Diabetes Care 2011;34:S11-S61.

(84) Coustan DR. Gestational diabetes mellitus. Clin Chem 2013;59:1310-21.

(85) Benhalima K, Hanssens M, Devlieger R, Verhaeghe J, Mathieu C. Analysis of Pregnancy Outcomes Using the New IADPSG Recommendation Compared with the Carpenter and Coustan Criteria in an Area with a Low Prevalence of Gestational Diabetes. Int J Endocrinol 2013;2013:248121.

(86) Coustan DR. Point: the American Diabetes Association and the International Association of Diabetes and Pregnancy study groups recommendations for diagnosing gestational diabetes should be used worldwide. Clin Chem 2012;58:1094-7.

(87) Langer O, Umans JG, Miodovnik M. The proposed GDM diagnostic criteria: a difference, to be a difference, must make a difference. J Matern Fetal Neonatal Med 2013;26:111-5.

(88) Long H, Cundy T. Establishing consensus in the diagnosis of gestational diabetes following HAPO: where do we stand?. Curr Diab Rep 2013;13:43-50.

(89) Murphy VE, Smith R, Giles WB, Clifton VL. Endocrine regulation of human fetal growth: the role of the mother, placenta, and fetus. Endocr Rev 2006;27:141-69.

(90) Walsh JM, McAuliffe FM. Prediction and prevention of the macrosomic fetus. Eur J Obstet Gynecol Reprod Biol 2012;162:125-30.

(91) Mayer C, Joseph KS. Fetal growth: a review of terms, concepts and issues relevant to obstetrics. Ultrasound Obstet Gynecol 2013;41:136-45.

(92) Henriksen T. The macrosomic fetus: a challenge in current obstetrics. Acta Obstet Gynecol Scand 2008;87:134-45.

(93) Godang K, Qvigstad E, Voldner N, Isaksen GA, Frøslie KF, Notthellen J, et al. Assessing body composition in healthy newborn infants: reliability of dual-energy x-ray absorptiometry. J Clin Densitom 2010;13:151-60.

(94) Norwegian Institute of Public Health. Medical birth registry of Norway. Available from http://www.fhi.no/helseregistre/medisinsk-fodselsregister

(95) Orskou J, Kesmodel U, Henriksen TB, Secher NJ. An increasing proportion of infants weigh more than 4000 grams at birth. Acta Obstet Gynecol Scand 2001;80:931-6.

(96) Meeuwisse G, Olausson PO. Increased birth weights in the Nordic countries. A growing proportion of neonates weigh more than four kilos. [Article in Swedish] Läkartidningen. 1998;95:5488-92.

(97) Lu Y, Zhang J, Lu X, Xi W, Li Z. Secular trends of macrosomia in southeast China, 1994-2005. BMC Public Health 2011;11:818.

(98) Chike-Obi U, David RJ, Coutinho R, Wu SY. Birth weight has increased over a generation. Am J Epidemiol 1996;144:563-9.

(99) Newsome CA, Shiell AW, Fall CH, Phillips DI, Shier R, Law CM. Is birth weight related to later glucose and insulin metabolism?--A systematic review. Diabet Med 2003;20:339-48.

(100) Fisher D, Baird J, Payne L, Lucas P, Kleijnen J, Roberts H, et al. Are infant size and growth related to burden of disease in adulthood? A systematic review of literature. Int J Epidemiol 2006;35:1196-210.

(101) Schellong K, Schulz S, Harder T, Plagemann A. Birth weight and long-term overweight risk: systematic review and a meta-analysis including 643,902 persons from 66 studies and 26 countries globally. PLoS One 2012;7:e47776.

(102) Walhovd KB, Fjell AM, Brown TT, Kuperman JM, Chung Y, Hagler DJ, Jr., et al. Long-term influence of normal variation in neonatal characteristics on human brain development. Proc Natl Acad Sci U S A 2012;109:20089-94.

(103) Armitage JA, Poston L, Taylor PD. Developmental origins of obesity and the metabolic syndrome: the role of maternal obesity. Front Horm Res 2008;36:73-84.

(104) Fall CH. The fetal and early life origins of adult disease. Indian Pediatr 2003;40:480-502.

(105) Schulz LC. The Dutch Hunger Winter and the developmental origins of health and disease. Proc Natl Acad Sci U S A 2010;107:16757-8.

(106) Barker DJ. Fetal origins of coronary heart disease. BMJ 1995;311:171-4.

(107) Plagemann A, Harder T, Schellong K, Schulz S, Stupin JH. Early postnatal life as a critical time window for determination of long-term metabolic health. Best Pract Res Clin Endocrinol Metab 2012;26:641-53.

(108) Drake AJ, Reynolds RM. Impact of maternal obesity on offspring obesity and cardiometabolic disease risk. Reproduction 2010 Sep;140:387-98.

(109) International Society for Developmental Origins of Health and Disease (DOHaD). Available from http://www.mrc.soton.ac.uk/dohad/index.asp

(110) Voldner N, Frøslie KF, Bo K, Haakstad L, Hoff C, Godang K, et al. Modifiable determinants of fetal macrosomia: role of lifestyle-related factors. Acta Obstet Gynecol Scand 2008;87:423-9.

(111) Roland MC, Friis CM, Voldner N, Godang K, Bollerslev J, Haugen G, et al. Fetal growth versus birthweight: the role of placenta versus other determinants. PLoS One 2012;7:e39324.

(112) Walker CL, Ho SM. Developmental reprogramming of cancer susceptibility. Nat Rev Cancer 2012;12:479-86.

(113) Hay WW, Jr. Recent observations on the regulation of fetal metabolism by glucose. J Physiol 2006;572(Pt 1):17-24.

(114) Voldner N. Modifiable determinants of newborn macrosomia and birth complications. PhD thesis. University of Oslo, Faculty of Medicine; 2010.

(115) Voldner N, Frøslie KF, Haakstad LAH, Bo K, Henriksen T. Birth complications, overweight, and physical inactivity. Acta Obstet Gynecol Scand 2009;88:550-5.

(116) Voldner N, Qvigstad E, Frøslie KF, Godang K, Henriksen T, Bollerslev J. Increased risk of macrosomia among overweight women with high gestational rise in fasting glucose. J Matern Fetal Neonatal Med 2010;23:74-81.

(117) Bergsjø P, Maltau J, Molne K, Nesheim B. Obstetrikk. [In Norwegian] Universitetsforlaget AS; 1998.

(119) Wu TL, Tsao KC, Chang CP, Li CN, Sun CF, Wu JT. Development of ELISA on microplate for serum C-reactive protein and establishment of age-dependent normal reference range. Clin Chim Acta 2002;322:163-8.

(120) Godang K, Frøslie KF, Henriksen T, Isaksen GA, Voldner N, Lekva T, et al. Umbilical cord levels of sclerostin, placental weight, and birth weight are predictors of total bone mineral content in neonates. Eur J Endocrinol 2013;168:371-8.

(121) Veierød M, Lydersen S, Laake, (eds.). Medical statistics in clinical and epidemiological research. Gyldendal Akademisk; 2012.

(122) Carey RG. Improving health care with control charts. Basic and advanced SPC methods and case studies. ASQ Quality Press; 2003.

(123) Meko D. Detrending. Available from http://www.ltrr.arizona.edu/~dmeko/notes_7.pdf

(124) Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Data Mining, Inference, and Prediction. Springer Series in Statistics; 2001.

(125) Di CZ, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel Functional Principal Component Analysis. Ann Appl Stat 2009;3:458-88.

(126) Crainiceanu C, Goldsmith AJ. Bayesian Functional Data Analysis Using WinBUGS. J Stat Softw 2010;32:1-33.

(127) Sturtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. J Stat Softw 2005;12:1-16.

(128) Ntzoufras I. Bayesian Modeling Using WinBUGS. Wiley; 2009.

(129) The R Foundation for Statistical Computing: R version 3.0.2 (2013-09-25). Available from http://www.r-project.org

(130) Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004;15:615-25.

(131) Flegal KM, Keyl PM, Nieto FJ. Differential Misclassification Arising from Nondifferential Errors in Exposure Measurement. Am J Epidemiol 1991;134:1233-44.

(132) Griebeler ML, Levis S, Beringer LM, Chacra W, Gomez-Marin O. Self-reported versus measured height and weight in Hispanic and non-Hispanic menopausal women. J Womens Health (Larchmt) 2011;20:599-604.

(133) Perez-Cueto FJ, Verbeke W. Reliability and validity of self-reported weight and height in Belgium. Nutr Hosp 2009;24:366-7.

(134) Hammami M, Koo WW, Hockman EM. Technical considerations for fan-beam dual-energy x-ray absorptiometry body composition measurements in pediatric studies. JPEN J Parenter Enteral Nutr 2004;28:328-33.

(135) Madan JC, Davis JM, Craig WY, Collins M, Allan W, Quinn R, et al. Maternal obesity and markers of inflammation in pregnancy. Cytokine 2009;47:61-4.

(136) Congdon P. Applied Bayesian Modelling. Wiley; 2003.

(137) McIntyre HD, Chang AM, Callaway LK, Cowley DM, Dyer AR, Radaelli T, et al. Hormonal and metabolic factors associated with variations in insulin sensitivity in human pregnancy. Diabetes Care 2010;33:356-60.

(138) Curry AE, Vogel I, Skogstrand K, Drews C, Schendel DE, Flanders WD, et al. Maternal plasma cytokines in early- and mid-gestation of normal human pregnancy and their association with maternal factors. J Reprod Immunol 2008;77:152-60.

(139) Friis CM, Paasche Roland MC, Godang K, Ueland T, Tanbo T, Bollerslev J, et al. Adiposity-related inflammation: effects of pregnancy. Obesity (Silver Spring) 2013;21:E124-E130.

(140) Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 2006;25:127-41.

(141) Frøslie KF, Roislien J, Laake P, Henriksen T, Qvigstad E, Veierod MB. Categorisation of continuous exposure variables revisited. A response to the Hyperglycaemia and Adverse Pregnancy Outcome (HAPO) Study. BMC Med Res Methodol 2010;10:103.

(142) Jenkins DJA, Woelver TMS, Jenkins AL. Fiber and other dietary factors affecting nutritient absorption and metabolism. In *Modern nutrition in Health and disease*, ninth edition. pp 679-698. Lippincott Williams & Wilkins; 1999.

(143) Polonsky KS, Given BD, Hirsch LJ, Tillil H, Shapiro ET, Beebe C, et al. Abnormal Patterns of Insulin-Secretion in Non-Insulin-Dependent Diabetes-Mellitus. N Engl J Med 1988;318:1231-9.

(144) Freckmann G, Hagenlocher S, Baumstark A, Jendrike N, Gillen R, Rössner K, et al. Continuous glucose profiles in healthy subjects under everyday life conditions and after different meals. J Diabetes Sci Technol. 2007;1:695-703

(145) Sosenko JM, Palmer JP, Greenbaum CJ, Mahon J, Cowie C, Krischer JP, et al. Increasing the accuracy of oral glucose tolerance testing and extending its application to individuals with normal glucose tolerance for the prediction of type 1 diabetes - The Diabetes Prevention Trial-Type 1. Diabetes Care 2007;30:38-42.

(146) Ramachandran R, Gravenstein KS, Metter EJ, Egan JM, Ferrucci L, Chia CW. Selective Contribution of Regional Adiposity, Skeletal Muscle, and Adipokines to Glucose Disposal in Older Adults. J Am Geriatr Soc 2012;60:707-12.

(147) Weijers RNM, Bekedam DJ, Goldschmidt HMJ, Smulders YM. The clinical usefulness of glucose tolerance testing in gestational diabetes to predict early postpartum diabetes mellitus. Clin Chem Lab Med 2006;44:99-104.

(148) Steele R. Influences of Glucose Loading and of Injected Insulin on Hepatic Glucose Output. Ann N Y Acad Sci 1959;82:420-30.

(149) Turner RC, Holman RR, Matthews D, Hockaday TDR, Peto J. Insulin Deficiency and Insulin Resistance Interaction in Diabetes - Estimation of Their Relative Contribution by Feedback Analysis from Basal Plasma-Insulin and Glucose-Concentrations. Metabolism 1979;28:1086-96.

(150) Bergman RN, Phillips LS, Cobelli C. Physiologic Evaluation of Factors Controlling Glucose-Tolerance in Man - Measurement of Insulin Sensitivity and Beta-Cell Glucose Sensitivity from the Response to Intravenous Glucose. J Clin Invest 1981;68(6):1456-67.

(151) Wallace TM, Matthews DR. The assessment of insulin resistance in man. Diabet Med 2002;19:527-34.

(152) Juge-Aubry CE, Somm E, Chicheportiche R, Burger D, Pernin A, Cuenod-Pittet B, et al. Regulatory effects of interleukin (IL)-1, interferon-beta, and IL-4 on the production of IL-1 receptor antagonist by human adipose tissue. J Clin Endocrinol Metab 2004;89:2652-8.

(153) Radaelli T, Uvena-Celebrezze J, Minium J, Huston-Presley L, Catalano P, Hauguel-de MS. Maternal interleukin-6: marker of fetal growth and adiposity. J Soc Gynecol Investig 2006;13:53-7.

(154) Wong E, Freiberg M, Tracy R, Kuller L. Epidemiology of cytokines: the Women On the Move through Activity and Nutrition (WOMAN) Study. Am J Epidemiol 2008;168:443-53.

(155) Mesman I, Roseboom TJ, Bonsel GJ, Gemke RJ, van der Wal MF, Vrijkotte TG. Maternal pre-pregnancy body mass index explains infant's weight and BMI at 14 months: results from a multi-ethnic birth cohort study. Arch Dis Child 2009;94:587-95.

(156) Stuebe AM, Landon MB, Lai Y, Spong CY, Carpenter MW, Ramin SM, et al. Maternal BMI, glucose tolerance, and adverse pregnancy outcomes. Am J Obstet Gynecol 2012;207:62-7.

(157) Yasui T, Uemura H, Yamada M, Matsuzaki T, Tsuchiya N, Noguchi M, et al. Associations of interleukin-6 with interleukin-1beta, interleukin-8 and macrophage inflammatory protein-1beta in midlife women. Cytokine 2008;41:302-6.

(158) Kerssen A, de Valk HW, Visser GHA. Day-to-day glucose variability during pregnancy in women with Type 1 diabetes mellitus: Glucose profiles measured with the Continuous Glucose Monitoring System. BJOG 2004;111:919-24.

(159) World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hypergycemia. Report of a WHO/IDF consultation. World Health Organization; 2006.

(160) Cousins L, Rigg L, Hollingsworth D, Brink G, Aurand J, Yen SSC. 24-Hour Excursion and Diurnal Rhythm of Glucose, Insulin, and C-Peptide in Normal-Pregnancy. Am J Obstet Gynecol 1980;136:483-8.

(161) Li J, Kuang Y, Mason CC. Modeling the glucose-insulin regulatory system and ultradian insulin secretory oscillations with two explicit time delays. J Theor Biol 2006;242:722-35.

(162) Baidal DA, Faradji RN, Messinger S, Froud T, Monroy K, Ricordi C, et al. Early Metabolic Markers of Islet Allograft Dysfunction. Transplantation 2009;87:689-97.

(163) Sosenko JM, Palmer JP, Rafkin LE, Krischer JP, Cuthbertson D, Greenbaum CJ, et al. Trends of Earlier and Later Responses of C-peptide to Oral Glucose Challenges With Progression to Type 1 Diabetes in Diabetes Prevention Trial-Type 1 Participants. Diabetes Care 2010;33:620-5.

(164) Ferrara A, Weiss NS, Hedderson MM, Quesenberry CP, Selby JV, Ergas IJ, et al. Pregnancy plasma glucose levels exceeding the American Diabetes Association thresholds, but below the National Diabetes Data Group thresholds for gestational diabetes mellitus, are related to the risk of neonatal macrosomia, hypoglycaemia and hyperbilirubinaemia. Diabetologia 2007;50:298-306.

(165) McGowan CA, McAuliffe FM. The influence of maternal glycaemia and dietary glycaemic index on pregnancy outcome in healthy mothers. Br J Nutr 2010;104:153-9.

(166) Scholl TO, Sowers M, Chen X, Lenders C. Maternal glucose concentration influences fetal growth, gestation, and pregnancy complications. Am J Epidemiol 2001;154:514-20.

(167) Ong KK, Diderholm B, Salzano G, Wingate D, Hughes IA, MacDougall J, et al. Pregnancy insulin, glucose, and BMI contribute to birth outcomes in nondiabetic mothers. Diabetes Care 2008;31:2193-7.

(168) Jeffery AN, Voss LD, Metcalf BS, Wilkin TJ. The impact of pregnancy weight and glucose on the metabolic health of mother and child in the south west of the UK. Midwifery 2004;20:281-9.

(169) Jovanovicpeterson L, Peterson CM, Reed GF, Metzger BE, Mills JL, Knopp RH, et al. Maternal Postprandial Glucose-Levels and Infant Birth-Weight - the Diabetes in Early-Pregnancy Study. Am J Obstet Gynecol 1991;164:103-11.

(170) Beardsall K, Diderholm BMS, Dunger DB. Insulin and carbohydrate metabolism. Best Pract Res Clin Endocrinol Metab 2008;22:41-55.

(171) International Association of Diabetes and Pregnancy Study Groups Recommendations on the Diagnosis and Classification of Hyperglycemia in Pregnancy. Diabetes Care 2010;33:676-82.

(172) Brunborg H. Hvor mange innvandrere er det – og blir det – i Norge? [In Norwegian] Samfunnsspeilet 3/2013 Available from http://www.ssb.no/befolkning/artikler-og-publikasjoner/hvor-mange-innvandrere-er-det-og-blir-det-i-norge

(173) Crozier SR, Robinson SM, Borland SE, Godfrey KM, Cooper C, Inskip HM. Do women change their health behaviours in pregnancy? Findings from the Southampton Women's Survey. Paediatr Perinat Epidemiol 2009;23:446-53.

(174) Buhling KJ, Winkel T, Wolf C, Kurzidim B, Mahmoudi M, Wohlfarth K, et al. Optimal timing for postprandial glucose measurement in pregnant women with diabetes and a non-diabetic pregnant population evaluated by the Continuous Glucose Monitoring System (CGMS (R)). J Perinat Med 2005;33:125-31.

(175) Gumbiner B, Van CE, Beltz WF, Ditzler TM, Griver K, Polonsky KS, et al. Abnormalities of insulin pulsatility and glucose oscillations during meals in obese noninsulin-dependent diabetic patients: effects of weight reduction. J Clin Endocrinol Metab 1996;81:2061-8.

(176) Clapp JF, III. Effect of dietary carbohydrate on the glucose and insulin response to mixed caloric intake and exercise in both nonpregnant and pregnant women. Diabetes Care 1998;21 Suppl 2:B107-12.

(177) Dunstan DW, Kingwell BA, Larsen R, Healy GN, Cerin E, Hamilton MT, et al. Breaking up prolonged sitting reduces postprandial glucose and insulin responses. Diabetes Care 2012;35:976-83.

Appendix: Supplementary results

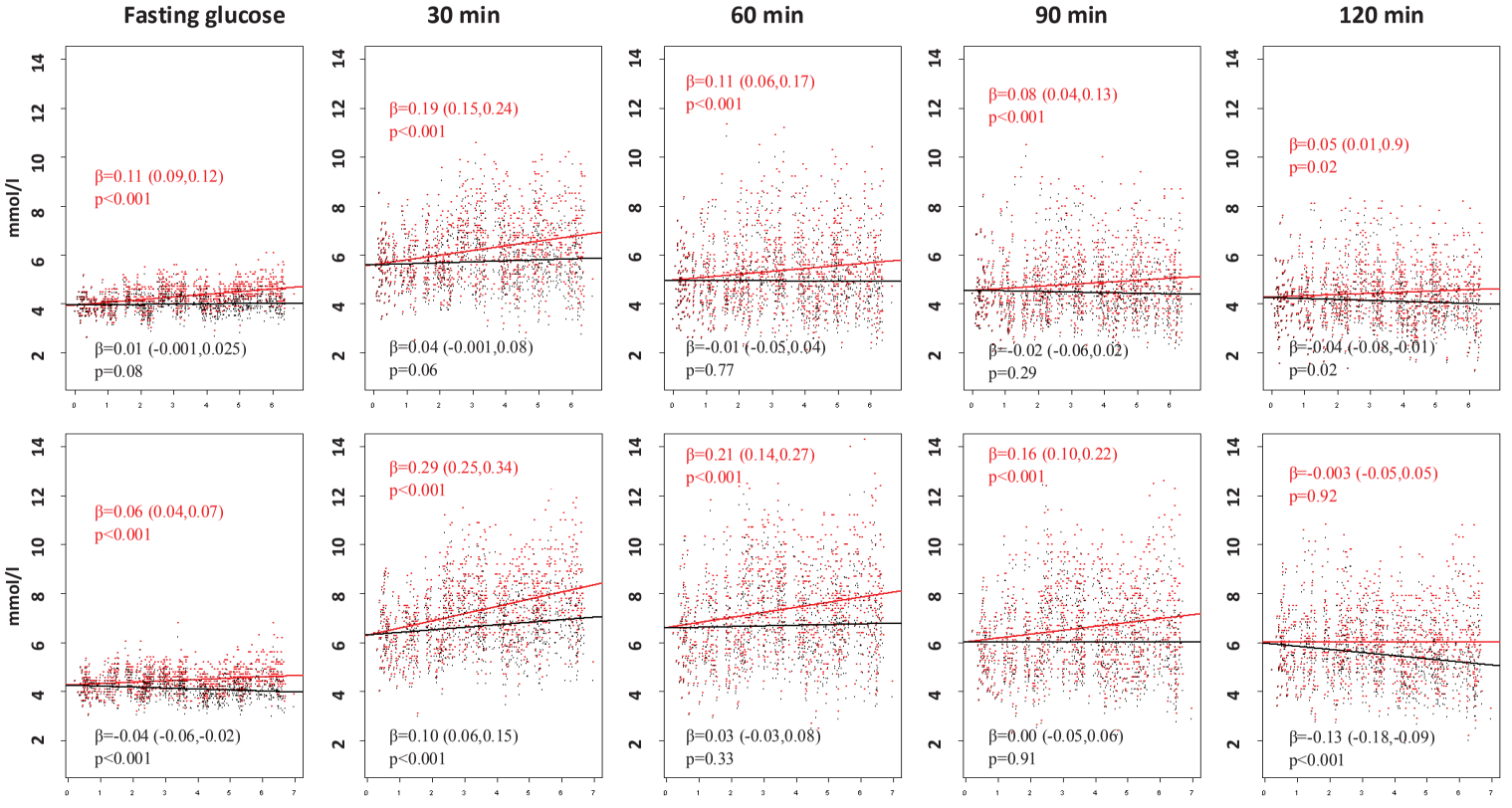


Figure 1. Time-dependent trends in measured (red) and de-trended (black) blood glucose at visit 1 (upper panel) and visit 3 (lower panel)

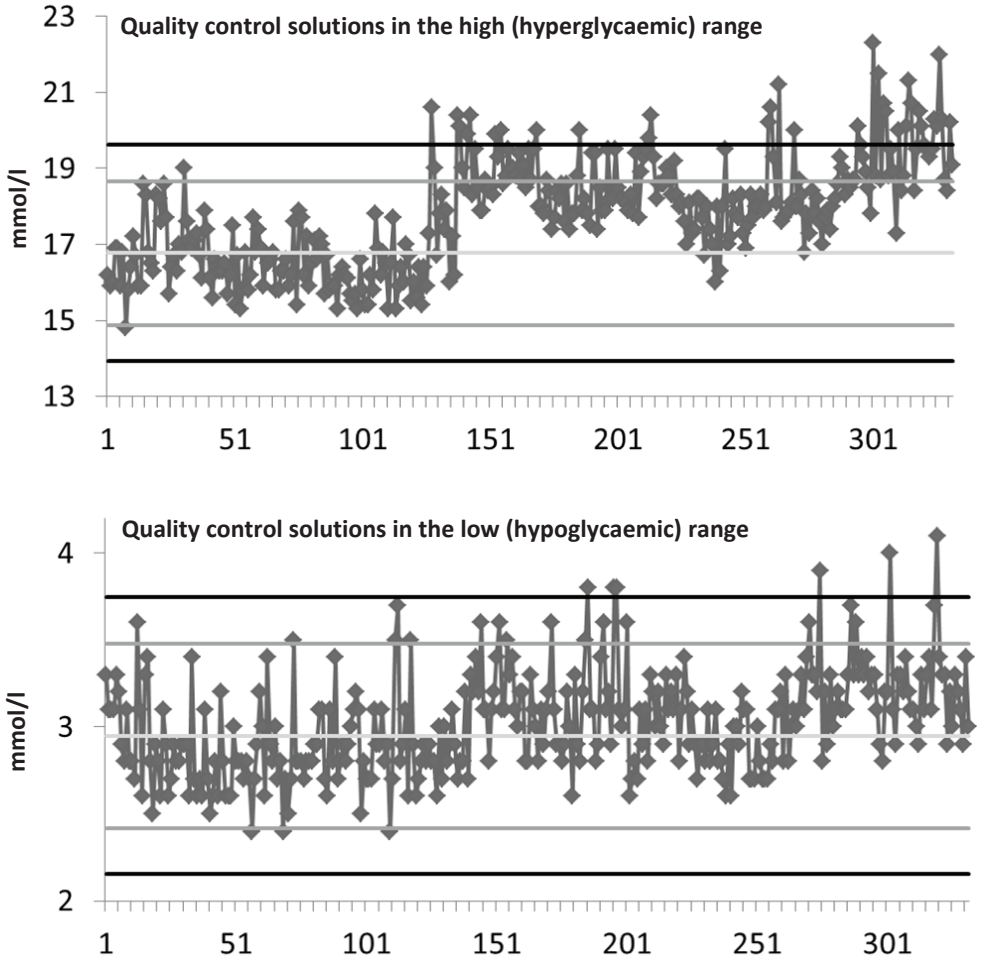


Figure 2. Control charts (X-charts) for the high and low quality control samples described in Article 1. The light grey horizontal line in each plot is the mean of the 30 first registrations, extended throughout the period. In each plot, the dark grey and black lines are the alarm limits, defined as mean ± 2SD of the first 30 consecutive data points, and the control limits, defined as mean ± 3SD of the first 30 consecutive data points, respectively.

I

Journal of Developmental Origins of Health and Disease, page 1 of 9.

doi:10.1017/S204017441000036X

ORIGINAL ARTICLE

The interleukins IL-6 and IL-1Ra: a mediating role in the associations between BMI and birth weight?

C. M. Friis[1,2]*†, K. F. Frøslie[2,3]†, J. Røislien[3], N. Voldner[1], K. Godang[4], T. Ueland[4,5], J. Bollerslev[4,6], M. B. Veierød[3] and T. Henriksen[1,6]

[1]*Division of Obstetrics and Gynaecology, Oslo University Hospital, Oslo, Norway*
[2]*Division of Obstetrics and Gynaecology, National Resource Centre for Women's Health, Oslo University Hospital Rikshospitalet, Oslo, Norway*
[3]*Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway*
[4]*Section of Specialized Endocrinology, Department of Endocrinology, Oslo University Hospital Rikshospitalet, Oslo, Norway*
[5]*Research Institute for Internal Medicine, Oslo University Hospital Rikshospitalet, Oslo, Norway*
[6]*Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway*

The biological mechanisms in the association between maternal body mass index (BMI) and birth weight are not well understood, but are likely to involve maternal plasma glucose levels and nutrient transport across the placenta, both important modulators of fetal growth. Adipose tissue contributes to circulating levels of interleukins that may affect glucose metabolism and possibly also placental transport of nutrients. We investigated possible mediating roles of Interleukin 6 (IL-6) and Interleukin 1 Receptor antagonist (IL-1Ra) in 208 pregnant women. Known and hypothesized dependencies between BMI in early pregnancy and fasting glucose, IL-1Ra and IL-6 at gestational weeks 30–32, and birth weight were specified in a path diagram. Standardized regression coefficients, expressing direct, indirect and total effects, were estimated by Bayesian path analysis. Mean (S.D.) BMI was 24.9 kg/m^2 (4.2) and mean (S.D.) birth weight 3748 g (454). The total effect of BMI on birth weight was 0.24 (95% credibility interval (CrI) [0.12, 0.36]). The direct effect of IL-1Ra on birth weight was not statistically significant, but significant effects of BMI on IL-1Ra (0.61, 95% CrI [0.51, 0.72]), of IL-1Ra on fasting glucose (0.17, 95% CrI [0.01, 0.34]) and of fasting glucose on birth weight (0.14, 95% CrI [0.01, 0.27]) implied an indirect pathway from BMI via IL-1Ra on birth weight. Approximately 20% of the effect of BMI on birth weight was mediated through IL-1Ra. For IL-6, no such effects were found. Our results indicate that IL-1Ra may be a mediator in the association between BMI and birth weight.

Received 11 March 2010; Revised 10 May 2010; Accepted 10 June 2010

Keywords: birth weight, BMI, inflammation, path analysis, pregnancy

Introduction

Birth weight is a result of a complex interaction between maternal, placental and fetal factors. Of the maternal factors, maternal body mass index (BMI) is a strong, independent and modifiable predictor of birth weight and has been estimated to account for roughly 10–20% of the variance in birth weight.[1–4] While numerous studies have shown an association between maternal BMI and birth weight, fewer studies have addressed the issue of biological mediators in this association.[5,6] Considering the increasing prevalence of maternal obesity and the long-term implications of birth weight on later health and disease, an understanding of how excess fat exerts effects on birth weight is important.[7] There are at least two ways in which maternal BMI may affect fetal growth, by modifying nutrient availability or by modifying placental nutrient transport. Traditionally, the link between maternal BMI and birth weight has been attributed in large to maternal hyperglycemia and partly to other metabolic alterations associated with obesity, that is, changes in nutrient availability.[8,9] However, the fact that BMI remains a significant determinant of birth weight, after correcting for glucose in traditional regression analysis and also in studies of glucose-tolerant women, indicates that other factors associated with maternal obesity are likely to play a role in fetal growth.[9–12] Studies of non-pregnant populations as well as animal experiments suggest a role of adipose tissue-derived inflammatory factors like interleukins (IL-6 and IL-1Ra), tumor necrosis factor (TNF) and other adipocytokines as molecular links between excess adipose tissue and deranged glucose metabolism, including increased insulin resistance.[13–15] The few studies concerning inflammatory factors and insulin resistance in pregnancy indicate that at least some of the same mechanisms are present during gestation.[16]

Direct effects of adipose tissue-derived factors on placental properties have also been suggested. There is evidence of maternal obesity affecting placental size and inflammatory properties[17] and preliminary data suggest that placental nutrient transport capacity may be directly affected by interleukins like

*Address for correspondence: C. M. Friis, Division of Obstetrics and Gynaecology, National Resource Centre for Women's Health, Oslo University Hospital Rikshospitalet, Oslo, Norway.
(Email Camilla.Friis@ous-hf.no)

† Joint first authors.

IL-6.[18] Thus, adipokines may have a role as mediators in the associations between maternal fat mass, maternal glucose levels, placental properties and birth weight.

We studied whether the interleukins IL-6 and IL-1Ra had mediating roles in the association between BMI and birth weight. There are conflicting data about the association between TNF and BMI.[19] We therefore chose not to include TNF as a potential mediator. Known and hypothesized dependencies between the chosen variables were depicted in a path diagram. Effect sizes were estimated by path analysis, which is a method where several multiple regression equations are combined to obtain estimates of direct and indirect effects.[20] We used data from healthy pregnant women without infections, sampled from a Norwegian cohort study.[21] To our knowledge, previous studies have not used path analysis in testing for inflammatory factors as mediators in the association between maternal BMI and birth weight.

Methods

The present work was performed in a subsample of the STORK study.[21] STORK is a prospective cohort study of healthy women of Scandinavian heritage who registered for obstetric care at Oslo University hospital Rikshospitalet from 2002 to 2008 ($n = 1030$). Exclusion criteria were multiple pregnancies, known pre-gestational diabetes and severe chronic diseases (lung, cardiac, gastrointestinal or renal). The women were scheduled for four examinations at gestational weeks 14–16, 22–24, 30–32 and 36–38. Maternal height was measured at the first visit and weight at each visit. Fasting glucose was measured at weeks 14–16 and 30–32. Data on age, parity, educational level, smoking status and pregestational BMI were registered. Gestational age was based on ultrasound measures made at weeks 17–19. Data on preeclampsia and hypertension were obtained from hospital records. Birth weight was measured with a calibrated scale.

The present subsample included 240 women from the first part of the STORK cohort ($n = 553$), enrolled during the period 2002–2005 (Fig. 1). Inflammatory markers were obtained from fasting blood samples at all four visits. A subsample was chosen due to the limited resources for cytokine and other biochemical analyses. Placental insufficiency may be associated with inflammatory changes,[22,23] and the subsample was therefore restricted to women giving birth to a baby above the 10th birth weight percentile (2962 g) of the cohort. Stratified random sampling based on birth weight below or above 4200 g was used to ensure that women with macrosomic babies were included. Women with possible infections indicated by a C-reactive protein (CRP) value above 10 mg/l,[24–26] extreme values on IL-1Ra or IL-6 (values beyond 3 S.D. in log scale), or missing data of any variable in the path analysis were excluded from the analyses. The final study sample comprised 208 women (Fig. 1).

The study was approved by the Regional Committees for Medical Research Ethics and all women gave their written informed consent.

Blood sampling and biochemical measurements

The blood samples were drawn in the morning, between 0730 and 0830 after an overnight fast, and were obtained from vein puncture in tubes containing ethylenediaminetetraacetic acid (EDTA). Plasma glucose was measured immediately in EDTA blood by Accu Chek Glucose Test strips (Roche Diagnostics,

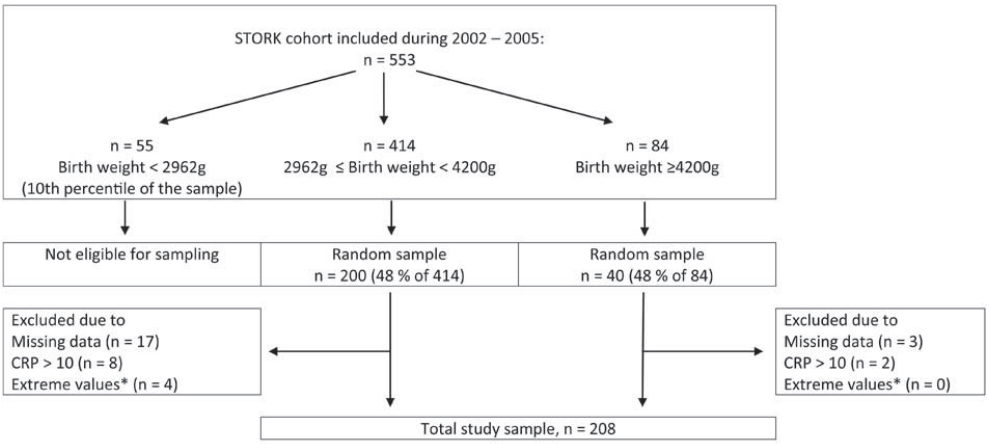


Fig. 1. Flow-chart showing the sampling procedure and the resulting study sample in this study.

Basel, Switzerland). The samples were immediately put on ice and plasma isolated, and stored at –80°C until analyzed. IL-6 (high sensitivity) and IL-1Ra were measured by ELISA using commercially available kits (BIOSOURCE, Invitrogen Corporation). CRP was measured as described by Wu *et al.*[24] All samples were measured in duplicate and serial samples from a given individual were analyzed at the same time to minimize the run-to-run variability. Intra- and inter-assay coefficients of variation were $<10\%$ for all assays.

Statistical methods

Descriptive statistics are presented as mean and standard deviation (S.D.), frequency and percentage (%) or as median and quartiles. Independent samples *t*-tests were used to compare the study sample ($n = 208$) and the remaining eligible women in the STORK cohort ($n = 258$, Fig. 1). Descriptive analyses and *t*-tests were performed by SPSS version 15.

Analytical methods in clinical research often rely on multiple regression models with one main outcome variable and explanatory variables treated on equal terms. Path analysis, in contrast, is a multivariable method based on a model with several linked regression equations.[20] Within this system of equations, some of the variables can be considered both as outcome variables and as explanatory variables. Path analysis is a form of structural equation modeling that requires that all hypothesized dependencies between the variables are specified in a model and depicted in a path diagram, prior to the analysis. Arrows in a path diagram represent dependencies between variables, and absence of an arrow between two variables indicates that these variables are considered statistically independent in the model. All direct and indirect relations among measured variables can be read off the path diagram.

Based on the literature, we constructed a path diagram for this study, which specified the hypothesized biological pathways between BMI in early pregnancy (weeks 14–16) and fasting glucose, IL-1Ra and IL-6 at weeks 30–32 (Fig. 2). The effect of BMI on birth weight was decomposed into a direct effect and indirect effects. The indirect pathways were hypothesized to be mediated by fasting glucose, the interleukins IL-1Ra or IL-6 or a combination of these, whereas the direct pathway incorporated other biological mechanisms and indirect pathways than those considered in this study. The path diagram formed the basis for the path analysis, in which we obtained direct and indirect effect estimates by combining four linear regression equations; one with birth weight as the outcome, one with fasting glucose as the outcome, one with IL-1Ra as the outcome and one with IL-6 as the outcome. In order to assess the validity of the underlying assumption of linearity in the separate linear regressions, generalized additive models[27] were used previous to the path analysis. All variables were entered into the analysis as standardized variables, and hence all effect estimates presented are standardized regression coefficients. Indirect effects can then be found by multiplication of the regression coefficients along a given path, and total effects can be found by summing all direct and indirect effects between two variables.

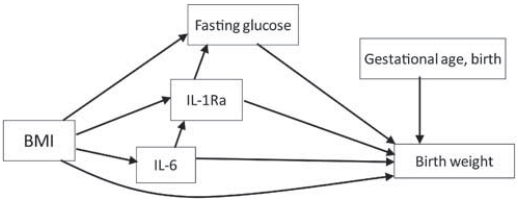


Fig. 2. Path diagram showing a decomposition of the hypothesized effect of maternal BMI on birth weight. The indirect pathways between BMI and birth weight were hypothesized to be mediated by fasting glucose (nutrient availability), the interleukins IL-1Ra or IL-6 or a combination of these. BMI was measured in the first trimester, fasting glucose and interleukins at weeks 30–32 and gestational age at birth. Arrows represent dependencies between variables. Absence of an arrow between two variables indicates that the variables are considered to be statistically independent in the model.

Path analyses were performed using Bayesian estimation procedures,[28] with the R2WinBUGS package,[29] that runs WinBUGS[30] from the statistical software R. Bayesian estimation gives estimates of regression coefficients and corresponding credibility intervals (CrIs), which are comparable with frequentistic confidence intervals. Considerations of statistical significance were based on the coverage of the credibility intervals. Comparison of models was carried out by the deviance information criterion (DIC); lower numbers of DIC are preferable.[31] Further details of the Bayesian model specification and model fitting can be found in Appendix A.

Results

Characteristics of the study sample are shown in Table 1. The study sample was not statistically different from those not selected to the present substudy ($0.08 \leqslant P \leqslant 0.95$), except from a significantly lower gestational age at visit 3 in the study sample (means 30.8 and 31.0 weeks, respectively, $P = 0.05$). Gestational age at birth ranged from 37.0 to 42.1 weeks. No significant bivariate correlations between gestational age at birth and BMI, fasting glucose, IL-1Ra or IL-6 were found in our study sample (results not shown).

Path analysis estimates of the model shown in Fig. 2, expressed as standardized regression coefficients and CrIs, are shown in Table 2. Gestational age at birth, maternal BMI in the early second trimester and fasting glucose in late pregnancy all had significant direct effects on birth weight. The strongest of these, with a standardized regression coefficient of 0.40 (95% CrI [0.28, 0.52]), was found for gestational age. This value implies that an increase of 1 S.D. in gestational age (1.2 weeks) gives a mean increase of 0.40 S.D. in birth weight (182 g). An alternative interpretation is that gestational age

Table 1. *Sample characteristics*

Sample characteristic	Study sample ($n = 208$[a])	STORK cohort BW ⩾ 2962 g, not selected for present substudy ($n = 258$[b])
Pre-gestational		
BMI (kg/m^2; self-reported)	23.6 (3.7)	23.7 (3.9)
Para 0	98 (47%)	136 (53%)
University education	172 (83%)	219 (85%)
Visit 1 (weeks 14–16)		
Gestational age (weeks)	15.7 (1.4)	15.7 (1.4)
Daily smokers[c]	5 (2%)	11 (4%)
Age of mother (years)	31.3 (3.9)	31.2 (4.2)
BMI (kg/m^2)	24.9 (4.2)	25.0 (4.2)
Fasting glucose (mmol/l)	4.2 (0.5)	4.2 (0.5)
Visit 3 (weeks 30–32)		
Gestational age (weeks)	30.8 (1.2)	31.0 (1.1)
Fasting glucose (mmol/l)	4.4 (0.5)	4.5 (0.5)
IL-1Ra (pg/ml; median) [Q_1,Q_3]	165 [136, 212]	NA
IL-6 (pg/ml; median) [Q_1,Q_3]	0.18 [0.10, 0.33]	NA
Preeclampsia	4 (2%)	10 (4%)
Hypertension	4 (2%)	4 (2%)
Birth		
Gestational age (weeks)	39.8 (1.2)	39.9 (1.4)
BW (g)	3748 (454)	3730 (451)
Boys	114 (55%)	138 (53%)

BW, birth weight; BMI, body mass index.

Selected characteristics of the study sample ($n = 208$) and 258 women from the STORK cohort who were not selected for this study. The numbers are mean (S.D.) or frequency (%) unless otherwise stated.

[a] Complete data on BMI at visit 1, fasting glucose, IL-1Ra and IL-6 at visit 3 and BW and gestational age at birth. Other numbers may vary due to missing values.

[b] Complete data on BW and gestational age at birth. Other numbers may vary due to missing values.

[c] More than 1 cigarette/day.

Table 2. *Path analysis*

		Standardized regression coefficients		
Outcome variable[a]	Effect	B	95% CrI	
BW	Direct effect gestational age → BW	0.40	0.28	0.52
	Direct effect BMI → BW	0.16	0.00	0.32
	Direct effect glucose → BW	0.14	0.01	0.27
	Direct effect IL-1Ra → BW	0.06	−0.10	0.21
	Direct effect IL-6 → BW	−0.02	−0.14	0.11
Fasting glucose	Direct effect BMI → fasting glucose	0.22	0.05	0.39
	Direct effect IL-1Ra → fasting glucose	0.17	0.01	0.34
IL-1Ra	Direct effect BMI → IL-1Ra	0.61	0.51	0.72
	Direct effect IL-6 → IL-1Ra	0.10	−0.01	0.21
IL-6	Direct effect BMI → IL-6	0.17	0.03	0.31

BW, birth weight; BMI, body mass index; CrI, credibility intervals.

Results from the path analysis illustrated in Figure 2. The table shows direct effects with corresponding CrI for the paths depicted in the figure, as well as the total effect of BMI on BW.

[a] BMI was measured in first trimester, fasting glucose and interleukins at weeks 30–32, and gestational age at birth.

accounts for 40% of the total variation in birth weight. The estimated direct effects for BMI and fasting glucose were 0.16 (95% CrI [0.00, 0.32]) and 0.14 (95% CrI [0.01, 0.27]), respectively.

The estimated direct effect of BMI on fasting glucose was also significant (0.22, 95% CrI [0.05, 0.39]), implying an indirect effect of BMI on birth weight mediated through glucose. An estimate of this indirect effect can be calculated from the direct effects of BMI on fasting glucose and fasting glucose on birth weight: $0.22 \cdot 0.14 = 0.03$.

There was no significant direct effect of IL-1Ra on birth weight. The effect of IL-1Ra on fasting glucose, however, was significant (0.17, 95% CrI [0.01, 0.34]), implying an indirect effect of IL-1Ra on birth weight. An estimate of this indirect effect can be calculated from the direct effects involved: $0.17 \cdot 0.14 = 0.02$. Hence, the effect of BMI on birth weight mediated through IL-1Ra was split into one path involving IL-1Ra only (estimated effect $0.61 \cdot 0.06 = 0.03$) and one path via IL-1Ra and fasting glucose (estimated effect $0.61 \cdot 0.17 \cdot 0.14 = 0.02$). In total, the estimated effect of BMI involving IL-1Ra was 0.05 (95% CrI [−0.05, 0.15]).

No significant direct effect of IL-6 on birth weight was found. In addition, the effect of IL-6 on IL-1Ra was not significant, indicating neither direct nor indirect effect of IL-6 on birth weight. The direct effect of BMI on IL-6 was significant (0.17, 95% CrI [0.03, 0.31]), but the indirect effect of BMI via IL-6 on birth weight was negligible in comparison to the other effects estimated in the model (<0.01, calculations not shown).

The *total* effect of maternal BMI on birth weight was estimated to be 0.24 (95% CrI [0.12, 0.36]). The total effect is the sum of the direct effect (0.16) and indirect effects via glucose only, via IL-1Ra only, via IL-1Ra and glucose and via IL-6 (Fig. 2), calculated above to be 0.03, 0.03, 0.02 and <0.01, respectively. The decomposition of the total effect of BMI on birth weight is emphasized in Fig. 3. The figure shows the relative percentages of the total BMI effect, through the different pathways in the model. Approximately 20% (0.05/0.24) of the total BMI effect worked through paths involving IL-1Ra, whereas a negligible percentage ($<1\%$) involved IL-6. Approximately 13% (0.03/0.24) of the effect worked through glucose without involving IL-1Ra. The remaining 67% (0.16/0.24) of the BMI effect represent effects not explained by variables or structures in our model.

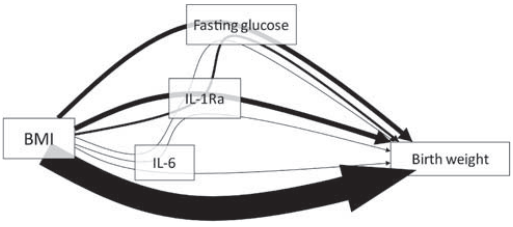


Fig. 3. The figure visualizes the decomposition of the total effect of maternal BMI on birth weight. The total effect is the sum of all arrows, that is, the direct and indirect effects. The arrow widths represent the relative proportions of the total effect through a specific pathway.

Considering the above results, a reduced model without IL-6 was formulated and the corresponding effects estimated (results not shown). DIC decreased substantially (from 2193 to 1603). The large reduction was mostly attributable to the weak association between BMI and IL-6, as the predictive capabilities of the model as a whole improves when leaving IL-6 out of the model. The effect of IL-1Ra on fasting glucose was not affected by the changes in the model. The direct effect of IL-1Ra on birth weight was still not significant, and the proportion of the total effect of BMI mediated through IL-1Ra was still approximately 20%.

Discussion

There are numerous reports on the effect of maternal BMI on birth weight,[4,6,9,11,21] but the biological mechanisms behind this association still remain to be elucidated. We have studied the possible mediating role of interleukins (IL-6 and IL-1Ra) in the association between BMI and birth weight. Path analysis indicated a mediating role of IL-1Ra, but less impact of IL-6.

The use of path diagrams and analysis of structural models is expanding in the field of epidemiology, including studies of pregnancy outcome,[6,32–34] and there is a need for biological understanding.[9,35] The crucial task in path analysis is to formulate a plausible path diagram based on existing evidence and current biological concepts. Maternal BMI may modify both nutrient availability and nutrient transport. Maternal BMI is a strong determinant of glucose plasma levels, and might thus indirectly affect birth weight through increasing nutrient availability for fetal growth.[36] We chose to include fasting glucose in the path diagram as some studies, including ours, indicate fasting glucose to correlate more strongly with both BMI and birth weight than the 2 h glucose value.[9,12,21,37,38] Late gestation glucose levels were included in the diagram, that is, when fetal growth is at its maximum.

In recent years, a growing body of literature has established a relation between BMI and low-grade systemic inflammation.[39] There is increasing evidence that the same relation is present during pregnancy.[40–42] Adipose tissue-derived inflammatory factors, including interleukins, have received considerable attention as potential mediators in the link between excess fat and the dysregulation of glucose metabolism including increased insulin resistance in obesity.[39,43] Our selection of potential inflammatory mediators was based on several considerations. In general, data from the non-pregnant population indicate that IL-6 and IL-1Ra are both central interleukins and interact with each other.[44] They are upstream markers of inflammation,[45] which are both elevated in obesity and have been implicated in glucose regulation in

epidemiological and experimental studies.[46–49] Furthermore, IL-6 has been found to be elevated in obese pregnant women.[40] Data are lacking for IL-1Ra and maternal obesity, but IL-1Ra is one of the most consistent markers of obesity in the non-pregnant population.[50]

Fetal growth is also dependent on nutrient transport across the placenta. Maternal obesity has been found to affect placental size, structure and function.[17] Inflammatory molecules do not easily pass the placenta and maternal inflammation does not seem to be associated with umbilical cord inflammation.[51,52] Thus, an effect of adipokines on birth weight will expectedly work through altered placental transport capacity or function. Indeed, preliminary data show an effect on transport proteins in the placenta after exposure to IL-6.[18] IL-6 may also act on the placenta and regulate fetal growth through upregulating leptin, which in turn regulates placental growth, nutrient transfer and fetal fat accretion.[53] Finally, experimental studies have demonstrated increased litter fat mass after prenatal exposure to IL-6[54] and one study has linked maternal IL-6 levels directly to neonatal fat mass.[55] Thus, the literature suggests a biological role for interleukins in the association between maternal BMI and birth weight.

The hypothesized associations were partly confirmed in our data. About 20% of the effect of BMI on birth weight was mediated through paths involving IL-1Ra. However, IL-1Ra is a dual marker; it is an anti-inflammatory cytokine, binding to IL-1 receptor without inducing an effect, but at the same time reflects an activation of the IL-1 system and is also a marker of inflammation in general.[50,56] Based on this, we cannot rule out that the measured effect of IL-1Ra reflects the action of IL-1β. This emphasizes the need for experimental studies to assess molecular mechanisms and also emphasizes the importance of interpreting results from observational studies with caution concerning causality. However, this result indicates a substantial role for the interleukin 1-system in the deranged glucose metabolism associated with higher maternal BMI during pregnancy and consequently an important role for interleukins as mediators between maternal fat mass, glucose and birth weight. We did not find significant direct effects of the interleukins on birth weight. There might be several explanations for this finding. It may be that interleukins like IL-6 and IL-1Ra do not play an important role in regulating fetal growth through changing placental properties. The result may however also be due to the fact that cytokines display pleiotrophic effects and show considerable biological variation. We chose two markers as representative of the inflammatory status in obese women, being aware that other markers may be important as mediators in the association between BMI and birth weight. In addition, effects of cytokines on birth weight are probably not an effect of a single mediator, but rather the result of the interactions of several and in combinations.[57,58] Therefore, we cannot rule out that cytokines in combinations may have a direct effect on placental properties and birth weight even if we were not able to find such an effect.

We recognize that the hypothesized path diagram (Fig. 1), in which maternal BMI leads to increased inflammation with secondary downstream effects on glucose regulation and fetal growth, has limitations. Integrative physiology is much more complex than reflected in this simplified model. For example, fetal growth relies primarily on glucose as an energy substrate; however lipids and amino acids are also nutrient substrates for fetal growth.[59] An expansion of the model to include lipids would be interesting.

Nevertheless, the analyses based on our simplified path diagram support the notion that inflammatory mediators are involved in the association between maternal BMI and birth weight.

BMI and birth weight are both surrogate markers of fat mass. We used BMI at early gestation in the path diagram, as maternal BMI and fat mass have been shown to correlate more strongly in early than in late pregnancy.[60] Glucose and inflammatory markers were analyzed at gestational weeks 30–32 but there is evidence that the inflammatory and metabolic derangements associated with pregravid maternal obesity are sustained throughout pregnancy.[41] The use of birth weight as a marker of fetal growth might explain why our results were not in accordance with the previously reported association between maternal IL-6 and prenatal growth reflected by neonatal fat mass.[55]

Bayesian methods have been used to a limited extent in clinical research, but the WinBUGS software has made Bayesian methods available.[28,61] Traditional frequentistic analyses are based on normality assumptions and central limit theory, whereas the WinBUGS analyses are based on prior assumptions and simulation techniques. For both approaches, the linearity of the regression equations should be explored.[27] Frequentistic path analysis is sensitive to violations of normality assumptions in small samples and non-linearity or combinations of different types of variables can be difficult to handle. In Bayesian models, in contrast, non-normality and non-linearity are more easily dealt with. Such methods are also flexible with respect to several types of variables.[28,61] In studies of complex biological mechanisms, the samples will typically be small due to the costs and restraints in collecting the data. In addition, inflammatory biological markers tend to be skewed, and sometimes display non-linear relations.[16,62] Therefore, Bayesian methods represent a valuable tool in such studies.

The representativity of the STORK cohort, considering voluntarily participation and a closer follow-up than in usual obstetric care in Norway, has been described earlier.[63] As our study focused on general physiological mechanisms, presumably similar in all healthy pregnant women, neither close follow-up nor self-selection effects would be likely to affect our results substantially.

We wanted to avoid confounding from other biological processes than those studied. Women with low infant birth weight were not included in this study because fetal growth restriction may be associated with placental inflammatory changes.[22,23] No formal definition of fetal growth restriction exists, and the use of the 10th birth weight percentile as an exclusion criterion was in

accordance with a pragmatic tradition. We used a CRP value above 10 mg/l to exclude women with possible infections,[24–26] thereby adjusting for confounding caused by infections. Unmeasured lifestyle factors (like diet or physical activity), genetic factors or biological factors might confound with the relations we studied, and thereby either attenuate or increase effects. However, it is hard to tell in what direction the effect sizes would be affected.[64] Gestational age at birth was modeled as a potential confounder of the direct effects on birth weight, but not on the causal pathway to birth weight. Our estimates would be strongly affected if this was the case.[65,66] However, no significant bivariate correlations between gestational age and the other variables were found in our study sample.

This study is based on a relatively large sample with measurements of inflammatory markers, which is a strength due to substantial biological variance of such markers.[57] Furthermore, moderate effect estimates were anticipated. As a basis for comparison, maternal BMI, one of the major determinants of birth weight, accounts for approximately 10–20% of the variation in birth weight.[1–4] Our result was similar, but only borderline significant, possibly due to the homogeneity of our study sample and a lack of power to detect small effects. Based on these considerations, we reported our model with all the original arrows present, although not all the direct effects were significant. Correspondingly, indirect effects were estimated with significant and non-significant direct effects included.

We conclude that the results of our study, combining current biological concepts and empirical data, suggest that adipose tissue-derived inflammatory factors may be mediators in the association between BMI and birth weight. Mechanisms like metabolic pathways are complex, yet simplified models like the current one may still be useful.

Acknowledgments

We thank Petter Laake, professor in statistics, for his contributions to our discussions on path analysis.

Statement of Interest

The study was supported by the Norwegian Health Association, the National Resource Centre for Women's Health, Division of Obstetrics and Gynaecology, Oslo University Hospital Rikshospitalet, Norway and the Faculty of Medicine, Thematic Research Area, University of Oslo, Norway.

References

1. Ehrenberg HM, Mercer BM, Catalano PM. The influence of obesity and diabetes on the prevalence of macrosomia. *Am J Obstet Gynecol*. 2004; 191, 964–968.
2. Mesman I, Roseboom TJ, Bonsel GJ, *et al*. Maternal pre-pregnancy body mass index explains infant's weight and BMI at 14 months: results from a multi-ethnic birth cohort study. *Arch Dis Child*. 2009; 94, 587–595.
3. Catalano PM, Drago NM, Amini SB. Factors affecting fetal growth and body composition. *Am J Obstet Gynecol*. 1995; 172, 1459–1463.
4. Catalano PM, Kirwan JP. Maternal factors that determine neonatal size and body fat. *Curr Diab Rep*. 2001; 1, 71–77.
5. Jansson N, Nilsfelt A, Gellerstedt M, *et al*. Maternal hormones linking maternal body mass index and dietary intake to birth weight. *Am J Clin Nutr*. 2008; 87, 1743–1749.
6. Fleten C, Stigum H, Magnus P, Nystad W. Exercise during pregnancy, maternal prepregnancy body mass index, and birth weight. *Obstet Gynecol*. 2010; 115, 331–337.
7. Armitage JA, Poston L, Taylor PD. Developmental origins of obesity and the metabolic syndrome: the role of maternal obesity. *Front Horm Res*. 2008; 36, 73–84.
8. King JC. Maternal obesity, metabolism, and pregnancy outcomes. *Annu Rev Nutr*. 2006; 26, 271–291.
9. HAPO Study Cooperative Research Group. Hyperglycaemia and Adverse Pregnancy Outcome (HAPO) Study: associations with maternal body mass index. *BJOG*. 2010; 117, 575–584.
10. Sacks DA, Liu AI, Wolde-Tsadik G, *et al*. What proportion of birth weight is attributable to maternal glucose among infants of diabetic women? *Am J Obstet Gynecol*. 2006; 194, 501–507.
11. Owens LA, O'Sullivan EP, Kirwan B, *et al*. ATLANTIC DIP: the impact of obesity on pregnancy outcome in glucose tolerant women. *Diabetes Care*. 2010; 33, 577–579.
12. Jensen DM, Damm P, Sorensen B, *et al*. Pregnancy outcome and prepregnancy body mass index in 2459 glucose-tolerant Danish women. *Am J Obstet Gynecol*. 2003; 189, 239–244.
13. Cai D, Yuan M, Frantz DF, *et al*. Local and systemic insulin resistance resulting from hepatic activation of IKK-beta and NF-kappaB. *Nat Med*. 2005; 11, 183–190.
14. Meier CA, Bobbioni E, Gabay C, *et al*. IL-1 receptor antagonist serum levels are increased in human obesity: a possible link to the resistance to leptin? *J Clin Endocrinol Metab*. 2002; 87, 1184–1188.
15. Bastard JP, Maachi M, Lagathu C, *et al*. Recent advances in the relationship between obesity, inflammation, and insulin resistance. *Eur Cytokine Netw*. 2006; 17, 4–12.
16. McIntyre HD, Chang AM, Callaway LK, *et al*. Hormonal and metabolic factors associated with variations in insulin sensitivity in human pregnancy. *Diabetes Care*. 2009.
17. Challier JC, Basu S, Bintein T, *et al*. Obesity in pregnancy stimulates macrophage accumulation and inflammation in the placenta. *Placenta*. 2008; 29, 274–281.
18. Lash GE, Ansari T, Bischof P, *et al*. IFPA meeting 2008 workshops report. *Placenta*. 2009; 30(Suppl A), S4–S14.
19. Madan JC, Davis JM, Craig WY, *et al*. Maternal obesity and markers of inflammation in pregnancy. *Cytokine*. 2009; 47, 61–64.
20. Armitage P, Colton T. Encyclopedia of biostatistics. In *Path Analysis* (ed. Bollen KA), Vol. 4. 1998; pp. 3280–3284. John Wiley & Sons Ltd, West Sussex, England.
21. Voldner N, Froslie KF, Bo K, *et al*. Modifiable determinants of fetal macrosomia: role of lifestyle-related factors. *Acta Obstet Gynecol Scand*. 2008; 87, 423–429.
22. Sitras V, Paulssen R, Leirvik J, Vartun A, Acharya G. Placental gene expression profile in intrauterine growth restriction due to placental insufficiency. *Reprod Sci*. 2009; 16, 701–711.

23. Ogge G, Romero R, Chaiworapongsa T, *et al*. Leukocytes of pregnant women with small-for-gestational age neonates have a different phenotypic and metabolic activity from those of women with preeclampsia. *J Matern Fetal Neonatal Med*. 2009.
24. Wu TL, Tsao KC, Chang CP, *et al*. Development of ELISA on microplate for serum C-reactive protein and establishment of age-dependent normal reference range. *Clin Chim Acta*. 2002; 322, 163–168.
25. Shine B, de Beer FC, Pepys MB. Solid phase radioimmunoassays for human C-reactive protein. *Clin Chim Acta*. 1981; 117, 13–23.
26. Belo L, Santos-Silva A, Rocha S, *et al*. Fluctuations in C-reactive protein concentration and neutrophil activation during normal human pregnancy. *Eur J Obstet Gynecol Reprod Biol*. 2005; 123, 46–51.
27. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, 2001. Springer Series in Statistics, Canada.
28. Ntzoufras I. *Bayesian Modeling Using WinBUGS*, 2009. Wiley, New Jersey.
29. Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R. *J Stat Softw*. 2005; 12, 1–16.
30. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput*. 2000; 10, 325–337.
31. Spiegelhalter DJ, Best NG, Carlin BR, van der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Series B-Stat Methodol*. 2002; 64, 583–616.
32. Fields SJ, Livshits G, Sirotta L, Merlob P. Path analysis of risk factors leading to premature birth. *Am J Hum Biol*. 1996; 8, 433–443.
33. Sulkes J, Fields S, Gabbay U, Hod M, Merlob P. Path analysis on the risk of mortality in very low birth weight infants. *Eur J Epidemiol*. 2000; 16, 337–341.
34. Gamborg M, Andersen PK, Baker JL, *et al*. Life course path analysis of birth weight, childhood growth, and adult systolic blood pressure. *Am J Epidemiol*. 2009; 169, 1167–1178.
35. Factor-Litvak P, Sher A. Invited commentary: coming out of the box. *Am J Epidemiol*. 2009; 169, 1179–1181.
36. Clausen T, Burski TK, Oyen N, *et al*. Maternal anthropometric and metabolic factors in the first half of pregnancy and risk of neonatal macrosomia in term pregnancies. A prospective study. *Eur J Endocrinol*. 2005; J, 887–894.
37. Ben-Haroush A, Hadar E, Chen R, Hod M, Yogev Y. Maternal obesity is a major risk factor for large-for-gestational-infants in pregnancies complicated by gestational diabetes. *Arch Gynecol Obstet*. 2009; 279, 539–543.
38. Jensen DM, Damm P, Sorensen B, *et al*. Clinical impact of mild carbohydrate intolerance in pregnancy: a study of 2904 nondiabetic Danish women with risk factors for gestational diabetes mellitus. *Am J Obstet Gynecol*. 2001; 185, 413–419.
39. Das UN. Is obesity an inflammatory condition? *Nutrition*. 2001; 17, 953–966.
40. Ramsay JE, Ferrell WR, Crawford L, *et al*. Maternal obesity is associated with dysregulation of metabolic, vascular, and inflammatory pathways. *J Clin Endocrinol Metab*. 2002; 87, 4231–4237.
41. Stewart FM, Freeman DJ, Ramsay JE, *et al*. Longitudinal assessment of maternal endothelial function and markers of inflammation and placental function throughout pregnancy in lean and obese mothers. *J Clin Endocrinol Metab*. 2007; 92, 969–975.
42. Retnakaran R, Hanley AJ, Raif N, *et al*. C-reactive protein and gestational diabetes: the central role of maternal obesity. *J Clin Endocrinol Metab*. 2003; 88, 3507–3512.
43. Lee DE, Kehlenbrink S, Lee H, Hawkins M, Yudkin JS. Getting the message across: mechanisms of physiological cross talk by adipose tissue. *Am J Physiol – Endocrinol Metab*. 2009; 296, E1210–E1229.
44. Fattori E, Cappelletti M, Costa P, *et al*. Defective inflammatory response in interleukin 6-deficient mice. *J Exp Med*. 1994; 180, 1243–1250.
45. Gabay C, Kushner I. Acute-phase proteins and other systemic responses to inflammation. *N Engl J Med*. 1999; 340, 448–454.
46. Feve B, Bastard JP. The role of interleukins in insulin resistance and type 2 diabetes mellitus. *Nat Rev Endocrinol*. 2009; 5, 305–311.
47. Saltevo J, Laakso M, Jokelainen J, *et al*. Levels of adiponectin, C-reactive protein and interleukin-1 receptor antagonist are associated with insulin sensitivity: a population-based study. *Diabetes Metab Res Rev*. 2008; 24, 378–383.
48. Somm E, Cettour-Rose P, Asensio C, *et al*. Interleukin-1 receptor antagonist is upregulated during diet-induced obesity and regulates insulin sensitivity in rodents. *Diabetologia*. 2006; 49, 387–393.
49. Hauguel-de MS, Guerre-Millo M. The placenta cytokine network and inflammatory signals. *Placenta*. 2005.
50. Juge-Aubry CE, Somm E, Giusti V, *et al*. Adipose tissue is a major source of interleukin-1 receptor antagonist: upregulation in obesity and inflammation. *Diabetes*. 2003; 52, 1104–1110.
51. Aaltonen R, Heikkinen T, Hakala K, Laine K, Alanen A. Transfer of proinflammatory cytokines across term placenta. *Obstet Gynecol*. 2005; 106, 802–807.
52. Catalano PM, Presley L, Minium J, Hauguel-de MS. Fetuses of obese mothers develop insulin resistance in utero. *Diabetes Care*. 2009; 32, 1076–1080.
53. Nuamah MA, Yura S, Sagawa N, *et al*. Significant increase in maternal plasma leptin concentration in induced delivery: a possible contribution of pro-inflammatory cytokines to placental leptin secretion. *Endocr J*. 2004; 51, 177–187.
54. Dahlgren J, Nilsson C, Jennische E, *et al*. Prenatal cytokine exposure results in obesity and gender-specific programming. *Am J Physiol Endocrinol Metab*. 2001; 281, E326–E334.
55. Radaelli T, Uvena-Celebrezze J, Minium J, *et al*. Maternal interleukin-6: marker of fetal growth and adiposity. *J Soc Gynecol Investig*. 2006; 13, 53–57.
56. Juge-Aubry CE, Somm E, Chicheportiche R, *et al*. Regulatory effects of interleukin (IL)-1, interferon-beta, and IL-4 on the production of IL-1 receptor antagonist by human adipose tissue. *J Clin Endocrinol Metab*. 2004; 89, 2652–2658.
57. Wong E, Freiberg M, Tracy R, Kuller L. Epidemiology of cytokines: the Women On the Move through Activity and Nutrition (WOMAN) Study. *Am J Epidemiol*. 2008; 168, 443–453.
58. Yasui T, Uemura H, Yamada M, *et al*. Associations of interleukin-6 with interleukin-1beta, interleukin-8 and macrophage inflammatory protein-1beta in midlife women. *Cytokine*. 2008; 41, 302–306.
59. Schaefer-Graf UM, Graf K, Kulbacka I, *et al*. Maternal lipids as strong determinants of fetal environment and growth in pregnancies with gestational diabetes mellitus. *Diabetes Care*. 2008; 31, 1858–1863.

60. Sewell MF, Huston-Presley L, Amini SB, Catalano PM. Body mass index: a true indicator of body fat in obese gravidas. *J Reprod Med*. 2007; 52, 907–911.
61. Congdon P. *Applied Bayesian Modelling*, 2003. Wiley, England.
62. Curry AE, Vogel I, Skogstrand K, *et al*. Maternal plasma cytokines in early- and mid-gestation of normal human pregnancy and their association with maternal factors. *J Reprod Immunol*. 2008; 77, 152–160.
63. Voldner N . Modifiable determinants of newborn macrosomia and birth complications. Dissertation for the Degree of PhD 2010. Faculty of Medicine, University of Oslo.
64. Fewell Z, Davey SG, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol*. 2007; 166, 646–655.
65. Delbaere I, Vansteelandt S, De Bacquer D, *et al*. Should we adjust for gestational age when analysing birth weights? The use of z-scores revisited. *Hum Reprod*. 2007; 22, 2080–2083.
66. Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*. 2002; 155, 176–184.[†] Joint first authors

Appendix A

All path estimates presented in this paper are based on MCMC samples from the joint posterior distribution of the parameters given in the data. We used three parallel MCMC chains in our calculations, each based on 30,000 iterations from which the first 10,000 were discarded as a 'burn-in' to achieve convergence, and a thinning factor of five to avoid autocorrelation in the samples. Inference was based on the remaining 12,000 iterations. Convergence of the MCMC series was confirmed using several plots and diagnostics available in the coda-package,[29] including density plots, trace plots, autocorrelation plots, the Gelman-Rubin diagnostic and the Raftery-Lewis diagnostic.[28] Vague prior probability distributions were used for all parameters.[28] Different parameter specifications of the priors were tried to check for the influence of choice of priors. Computing code used to implement the models is available as supplementary material at the journal website (S1).

Supplementary material, Article I

```
##### Model1.bug #####

model {
       for (j in 1:J)
       {
       fvsd[j]       ~ dnorm(mu.fv[j], tau.fv)
       mu.fv[j]      <- a.0 + a.gest*gestsd[j] + a.bmi*bmisd[j] + a.g03*g03sd[j] +
                                a.il1ra3*il1ra3sd[j] + a.il63*il63sd[j]

       g03sd[j]      ~ dnorm(mu.g03[j], tau.g03)
       mu.g03[j]     <- b.0 + b.bmi*bmisd[j] + b.il1ra3*il1ra3sd[j]

       il1ra3sd[j]   ~ dnorm(mu.il1ra3[j], tau.il1ra3)
       mu.il1ra3[j] <- c.0 + c.bmi*bmisd[j] + c.il63*il63sd[j]

       il63sd[j]     ~ dnorm(mu.il63[j], tau.il63)
       mu.il63[j]    <- d.0 + d.bmi*bmisd[j]
       }

       a.0           ~ dnorm(0,0.01)
       a.gest        ~ dnorm(0,0.01)
       a.bmi         ~ dnorm(0,0.01)
       a.g03         ~ dnorm(0,0.01)
       a.il1ra3      ~ dnorm(0,0.01)
       a.il63        ~ dnorm(0,0.01)

       b.0           ~ dnorm(0,0.01)
       b.bmi         ~ dnorm(0,0.01)
       b.il1ra3      ~ dnorm(0,0.01)

       c.0           ~ dnorm(0,0.01)
       c.bmi         ~ dnorm(0,0.01)
       c.il63        ~ dnorm(0,0.01)

       d.0           ~ dnorm(0,0.01)
       d.bmi         ~ dnorm(0,0.01)

       tau.fv        ~ dgamma(0.5,0.5)
       sigma.fv      <- 1/tau.fv
       tau.g03       ~ dgamma(0.5,0.5)
       sigma.g03     <- 1/tau.g03
       tau.il1ra3    ~ dgamma(0.5,0.5)
       sigma.il1ra3 <- 1/tau.il1ra3
       tau.il63      ~ dgamma(0.5,0.5)
       sigma.il63    <- 1/tau.il63

       il1ra3tot     <- a.il1ra3 + b.il1ra3*a.g03
       il63tot       <- c.il63*b.il1ra3*a.g03 + c.il63*a.il1ra3 + a.il63

       bmitot        <- d.bmi*(c.il63*b.il1ra3*a.g03 + c.il63*a.il1ra3 + a.il63)+
                          c.bmi*(b.il1ra3*a.g03 + a.il1ra3 )+
                          b.bmi*a.g03 +
                          a.bmi
       bmiviainf     <- d.bmi*(c.il63*b.il1ra3*a.g03 + c.il63*a.il1ra3 + a.il63)+
                          c.bmi*(b.il1ra3*a.g03 + a.il1ra3 )
       }
```

```
##### Path_analysis.R #####

library(R2WinBUGS)
library("coda")

bwsd.frame   <- data.frame(cbind(fvsd,bmisd,gestsd,g03sd,il1ra3sd,il63sd))
attach(bwsd.frame)

J            <- length(fvsd)
data         <- list("J","fvsd","gestsd","bmisd","g03sd","il1ra3sd","il63sd")

inits        <- function(){
                  list( a.0          =rnorm(1,0,1),
                        a.gest       =rnorm(1,0,1),
                        a.bmi        =rnorm(1,0,1),
                        a.g03        =rnorm(1,0,1),
                        a.il1ra3     =rnorm(1,0,1),
                        a.il63       =rnorm(1,0,1),
                        b.0          =rnorm(1,0,1),
                        b.bmi        =rnorm(1,0,1),
                        b.il1ra3     =rnorm(1,0,1),
                        c.0          =rnorm(1,0,1),
                        c.bmi        =rnorm(1,0,1),
                        c.il63       =rnorm(1,0,1),
                        d.0          =rnorm(1,0,1),
                        d.bmi        =rnorm(1,0,1))}

Model.sim    <- bugs(data,inits,model.file="Model1.bug",
                  parameters.to.save=c("a.0","a.gest","a.bmi","a.g03","a.il1ra3",
                  "a.il63","b.0","b.bmi","b.il1ra3","c.0","c.bmi","c.il63",
                  "d.0","d.bmi","il1ra3tot","il63tot","bmitot","bmiviainf",
                  "sigma.fv","sigma.g03","sigma.il1ra3","sigma.il63"),
                  n.chains=3,
                  n.iter=30000,
                  n.burnin=10000,
                  n.thin=5,
#                 codaPkg=TRUE,       # For convergence diagnostics
#                 debug=TRUE,         # For log file display
                  bugs.directory="C:/Program Files/WinBUGS14")
```

II

III

Frøslie et al. BMC Medical Research Methodology 2013, 13:6
http://www.biomedcentral.com/1471-2288/13/6

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

Shape information from glucose curves: Functional data analysis compared with traditional summary measures

Kathrine Frey Frøslie[1,2*], Jo Røislien[1], Elisabeth Qvigstad[3], Kristin Godang[3], Jens Bollerslev[3,4], Nanna Voldner[5], Tore Henriksen[4,5] and Marit B Veierød[1]

Abstract

Background: Plasma glucose levels are important measures in medical care and research, and are often obtained from oral glucose tolerance tests (OGTT) with repeated measurements over 2–3 hours. It is common practice to use simple summary measures of OGTT curves. However, different OGTT curves can yield similar summary measures, and information of physiological or clinical interest may be lost. Our mean aim was to extract information inherent in the shape of OGTT glucose curves, compare it with the information from simple summary measures, and explore the clinical usefulness of such information.

Methods: OGTTs with five glucose measurements over two hours were recorded for 974 healthy pregnant women in their first trimester. For each woman, the five measurements were transformed into smooth OGTT glucose curves by functional data analysis (FDA), a collection of statistical methods developed specifically to analyse curve data. The essential modes of temporal variation between OGTT glucose curves were extracted by functional principal component analysis. The resultant functional principal component (FPC) scores were compared with commonly used simple summary measures: fasting and two-hour (2-h) values, area under the curve (AUC) and simple shape index (2-h minus 90-min values, or 90-min minus 60-min values). Clinical usefulness of FDA was explored by regression analyses of glucose tolerance later in pregnancy.

Results: Over 99% of the variation between individually fitted curves was expressed in the first three FPCs, interpreted physiologically as "general level" (FPC1), "time to peak" (FPC2) and "oscillations" (FPC3). FPC1 scores correlated strongly with AUC (r=0.999), but less with the other simple summary measures ($-0.42 \leq r \leq 0.79$). FPC2 scores gave shape information not captured by simple summary measures ($-0.12 \leq r \leq 0.40$). FPC2 scores, but not FPC1 nor the simple summary measures, discriminated between women who did and did not develop gestational diabetes later in pregnancy.

Conclusions: FDA of OGTT glucose curves in early pregnancy extracted shape information that was not identified by commonly used simple summary measures. This information discriminated between women with and without gestational diabetes later in pregnancy.

Keywords: Area under the curve, Curve shape, Functional data analysis, Functional principal component analysis, Gestational diabetes, Glucose curve, Glucose oscillations, Glucose variability, Oral glucose tolerance test, Pregnancy

* Correspondence: k.f.froslie@medisin.uio.no
[1]Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Boks 1122, Blindern 0317, Oslo, Norway
[2]Norwegian Resource Centre for Women's Health, Division of Obstetrics and Gynaecology, Oslo University Hospital, Rikshospitalet, Norway
Full list of author information is available at the end of the article

Background

Plasma glucose level is one of the most commonly used metabolic measures, both in research and in clinical settings [1-4]. In persons with normal glucose tolerance and metabolism, glucose levels rise after a dietary intake, and usually return to normal, postprandial levels after 2–3 hours [5,6]. For practical purposes, oral glucose tolerance test (OGTT) is used to define glucose tolerance [5,7,8]. Numerous studies have shown that high OGTT values are associated with an increased risk of adverse health outcomes [2-4,9], but there is no general agreement with respect to time points for glucose sampling during OGTT, cut-off values or test duration [1,2,4,10].

OGTT values are discrete, ordered measurements from an underlying, continuous process; i.e. an individual's glucose regulation. Temporal OGTT measurements are often used to illustrate the underlying glucose curves, but the information inherent in the shape of these curves has been the subject of few studies [11-14]. It is common practice to use simple summary measures, such as fasting value, two-hour (2-h) value or area under the curve (AUC) to obtain information about an individual's glucose tolerance. Simple summary measures are also frequently used in studies with continuous glucose monitoring [15,16]. To gain more information from OGTT glucose curves, simple shape summaries (shape indices), have been suggested [11-13]. However, different OGTT glucose curve trajectories can yield similar simple summary measures, and information of physiological or clinical interest may consequently be lost.

Functional data analysis (FDA) is a collection of statistical techniques specifically developed to analyse curve data [17-19]. When applying FDA, the entire curve is used as the basic unit of information, instead of the OGTT measurements at specific time points. FDA has been applied in some research disciplines during the last couple of decades, and has yielded novel insights of clinical importance in neuroscience [20], nephrology [21] and studies of gait [22,23]. An important FDA technique is functional principal component analysis (FPCA), which is used to extract the common temporal characteristics of a set of curves [18].

The main aim was to study the usefulness of FDA in the analysis of OGTT glucose curve trajectories. FDA, and in particular FPCA, was used to analyse OGTT data in a Norwegian prospective cohort study of healthy pregnant women [24]. We extracted temporal information from the shape of OGTT glucose curves and compared this to the information obtained from standard simple summary measures. By regression analyses we studied the OGTT glucose curves in relation to body mass index (BMI) categories in early pregnancy and gestational diabetes mellitus (GDM) later in pregnancy.

Methods

Participants and data

The STORK study is a prospective cohort of 1031 healthy pregnant women of Scandinavian heritage who registered for obstetric care at the Oslo University Hospital Rikshospitalet from 2001 to 2008 [25]. Exclusion criteria were multiple pregnancy, known history of type 1 or type 2 diabetes mellitus, and severe chronic diseases (pulmonary, cardiac, gastrointestinal, or renal). The overall aim of the STORK study was to gain insights into maternal metabolic syndrome and the determinants of foetal macrosomia [25]. Results of a 75 g OGTT, age, height and weight were recorded at inclusion at gestational weeks 14–16. Fifty-seven women (5.5%) with incomplete OGTT data were excluded, yielding a study sample of 974 women. During follow-up, 2-h glucose values at gestational weeks 30–32 were available for 930 (95%) women.

Venous blood samples were collected for OGTT in tubes containing Ethylenediaminetetraacetic acid (EDTA) between 07:30 and 08:30 after an overnight fast. Fasting glucose was measured immediately in a drop of fresh, whole EDTA blood, and further blood samples were taken every 30 minutes for 2 h, for a total of five OGTT measurements per woman. Glucose measurements were done by the Accu-Chek Sensor glucometer (Roche Diagnostics, Mannheim, Germany). Inter-assay coefficient of variation was <10%. Due to an unexpected increasing trend in fasting glucose values over the 7 years of participant recruitment, all glucose measurements were de-trended prior to the present analyses, as previously described in detail [26].

The study was approved by the Regional Committee for Medical Research Ethics, Southern Norway, Oslo, Norway (reference number S-01191), and performed according to the Declaration of Helsinki. All participating women provided written informed consent.

Data description

Descriptive statistics were mean, standard deviation (SD) and range, or frequency and percentage. The study sample and women with incomplete OGTT data were compared by two-sample t tests or χ^2 tests.

Functional data analysis

FDA is a common term for statistical techniques specifically developed for analysing curve data [17-19]. In FDA a temporal set of observations is transformed into a single, functional object, and statistical analysis is then performed on this continuous function, rather than on the original discrete data points. This makes it possible to extract information from the temporal process as a whole, instead of merely point-by-point. In a sample of curves, the mean curve is used descriptively, as in

traditional statistical analyses, and with proper modification, most standard statistical methods can be phrased in the framework of FDA. The principles of the analyses are explained hereafter, and technical details are given in the appendices.

Curve fitting

The five OGTT measurements for the 974 participating woman were converted into 974 continuous, smooth curves by subject-specific spline smoothing with B-splines basis functions [17,19] (Appendix A). These individually fitted curves formed the basis for the subsequent FDA.

Functional principal component analysis

FPCA was used to study the temporal variation in the 974 fitted curves. FPCA extracts a limited number of FPC curves that describe the temporal patterns associated with the largest proportions of the variation in the individual, fitted curves [17-19] (Appendix B). The FPC curves represent independent parts of the overall variability between the individual, fitted curves. The FPCA also yield individual FPC scores for each curve. The score variables are per definition independent, and the variation within the scores of an FPC quantifies the magnitude of the total variance explained by this FPC. A woman's FPC score for an FPC curve reflects how her individual curve trajectory corresponds to the general temporal feature expressed by this FPC curve. By FPCA it is thus possible to study how OGTT glucose curve trajectories vary from woman to woman. FPC curves are often illustrated by plots showing how an individual curve differs from the mean curve if the FPC scores are high or low, rather than plots of the FPC curves directly [17-19]. As in traditional principal component analysis, FPCs may be interpreted and labelled according to the information they exhibit, which in turn can be related to more conventional physiological or clinical theories.

Functional principal component scores vs simple summary measures

The Pearson correlation coefficient (r) was used to assess the associations between FPC scores, original glucose measurements and several simple summary measures of OGTT: fasting value, 2-h value, AUC and a simple shape index. We used the most cited simple shape index for OGTT [12], defined as the 2-h value minus the 90-min value for curves classified as "monophasic" or "biphasic", and the 90-min value minus the 60-min value for curves classified as "triphasic". The classification of curves, i.e. the determination of the number of phases within a curve involves an empirically chosen glucose threshold of 0.25 mmol/l [12]. Curves that did not meet the criteria for classification into mono-, bi- or triphasic were labelled "unclassified" and left out of the analyses.

Functional analysis of variance

The relation between BMI and simple summary measures of glucose values is well-known [27]. Functional analysis of variance (FANOVA), the functional counterpart of traditional analysis of variance (ANOVA), was used to analyse the effect of BMI on the shape of OGTT glucose curves [18], using the fitted curves as responses. The WHO classification for BMI was utilised (underweight (<18.5 kg/m^2), normal weight (18.5-25 kg/m^2, reference category), overweight (25–30 kg/m^2) and obese (≥30 kg/m^2) [27]) and BMI was entered as a categorical explanatory variable. The analysis was based on the shape of the mean curve in each BMI category, and the temporal differences between these curves (Appendix C). In FANOVA, the effect estimates are themselves curves over the same time span as the curves under study, i.e. OGTT glucose curves. Functional 95% confidence intervals (CIs) and p curves were obtained for the difference between two mean curves. The FANOVA also gives an overall p value for the difference between two BMI categories.

FANOVA vs ANOVA of simple summary measures

The simple summary measures described previously were compared across the BMI categories using traditional ANOVA, with Bonferroni corrected post hoc tests.

Curve shape information in regression analyses

There is an on-going discussion about the diagnostic criterion for GDM [28,29]. However, as a new international consensus has yet to be established, we have kept the GDM definition which at present is recommended by the WHO: a 2-h OGTT value of 7.8 mmol/l or higher [1]. Consequently, the 2-h value is important in current clinical practice. The impact of the curve shape in early pregnancy on glucose intolerance later in pregnancy, i.e. the 2-h value at gestational weeks 30–32, was assessed by regression analyses, using the FPC scores at gestational weeks 14–16 as explanatory variables.

To visualise the clinical usefulness of the curve shape information more clearly, and to account for potential non-linear relations between variables, the 2-h values at gestational weeks 30–32 were grouped into seven categories and multinomial logistic regression was performed [30] using this categorised variable as the response. The categories were based on the diagnostic criterion for GDM and on assessments of group size and percentiles in the sample: <3.27 (2.5th percentile), [3.27, 3.89) (2.5th-10th percentile), [3.89, 6.39) (10th-75th percentile; reference category), [6.39, 6.90) (75th-85th

percentile), [6.90, 7.8) (85th percentile to diagnostic cut-off for GDM) [7.8, 8.84) (GDM diagnosis to 98th percentile) and ≥8.84 mmol/l.

Five different models were fitted. Model 1 included BMI and the three independent FPC score variables from gestational weeks 14–16 as covariates, while models 2–5 included BMI and either the fasting value, the 2-h value, the AUC or the shape index, all from gestational weeks 14–16, as covariates. These simple measures were included one at a time in models 2–5, due to colinearity. Other covariates were not included in the models. It is beyond the scope of the article to build an extensive prediction model or to adjust for variables possibly on the causal pathway to the outcome. All covariates were continuous.

Software

FDA, i.e. curve fitting, FPCA and FANOVA, were performed using the fda package in R 2.13.0 [31]. The multinomial regression was done by the mlogit package in R 2.13.0 [31]. The R script is available as supplementary material [see Additional file 1]. All other analyses were performed in SPSS 19.

Results

Data description

Characteristics of the study sample at gestational weeks 14–16 are shown in Table 1. The women in the study sample were not significantly different from those with incomplete OGTT data ($0.11 \leq p \leq 0.94$). The number of women with a GDM diagnosis increased from 3 (0.3%) at gestational weeks 14–16 to 51 (5.5%) at gestational weeks 30–32 (Table 1).

Curve fitting

The individually fitted, smooth OGTT glucose curves at gestational weeks 14–16 showed large variations between the individual curves (Figure 1).

Functional principal component analysis

The essential modes of temporal variation between the fitted curves were extracted by FPCA (Figure 2). The first FPC (FPC1, Figure 2a) explained 88.1% of the variation between the fitted curves, the second FPC (FPC2, Figure 2b) 8.6% and the third FPC (FPC3, Figure 2c) 2.4%, respectively. The corresponding physiological interpretations were the general glucose level (FPC1, "general level"), the time to peak for glucose (FPC2,

Table 1 Sample characteristics

Characteristic	Study sample, *n*=974[a]		Excluded[b], *n*=57[a]	Total cohort, *n*=1031[a]
		Range		
Gestational weeks	15.8 (1.3)	12.1-22.0	16.0 (1.4)	15.8 (1.3)
Age	31 (4)	19-42	31 (4)	31 (4)
Para 0	517 (54%)		28 (50%)	545 (53%)
Daily smoker[c]	27 (3%)		1 (2%)	28 (3%)
Height (cm)	169 (6)	150-184	169 (6)	169 (6)
Weight (kg)	69.9 (12.0)	44.6-123.1	68.2 (12.5)	69.8 (12.0)
BMI (kg/m^2)	24.5 (3.9)	17.2-44.0	23.4 (3.8)	24.5 (3.9)
Birth weight[d] (g)	3588 (570)	600-5420	3554 (671)	3586 (576)
Blood glucose (mmol/l), first trimester				
Fasting	4.0 (0.4)	2.6-5.3		4.0 (0.4)
30 min	5.7 (1.2)	2.5-9.7		5.7 (1.2)
60 min	5.0 (1.4)	2.0-10.9		4.9 (1.4)
90 min	4.5 (1.2)	2.0-10.1		4.5 (1.2)
2 h	4.1 (1.1)	1.2-7.8		4.1 (1.1)
GDM[e]: 2-h value≥7.8 mmol/l	3 (0.3%)			3 (0.3%)
Blood glucose (mmol/l), third trimester				
2 h	5.5 (1.3)	1.9-10.3		5.5 (1.3)
GDM[e]: 2-h value≥7.8 mmol/l	51 (5.5%)			54 (5.5%)

Data are mean (SD) or frequency (%).
[a] Numbers may not add up to total due to missing data for some variables.
[b] Women excluded due to incomplete OGTT data.
[c] ≥1 cigarette/day.
[d] Birth weight of offspring.
[e] Gestational diabetes mellitus.

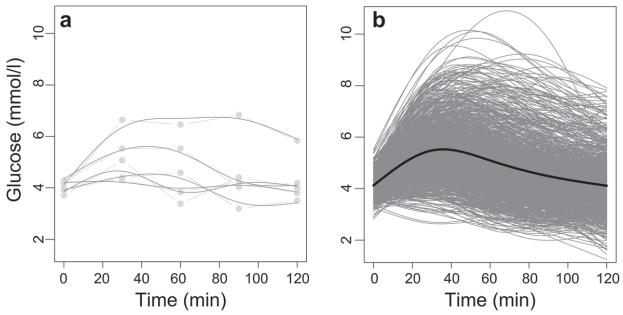


Figure 1 Observed OGTT data and individually fitted curves at gestational weeks 14–16. a shows the observed OGTT data (light grey) and individually fitted curves (dark grey) for the first five women in the study. The straight lines indicate measurements from the same woman. **b** shows the 974 individually fitted curves (grey) and the mean of these curves (black).

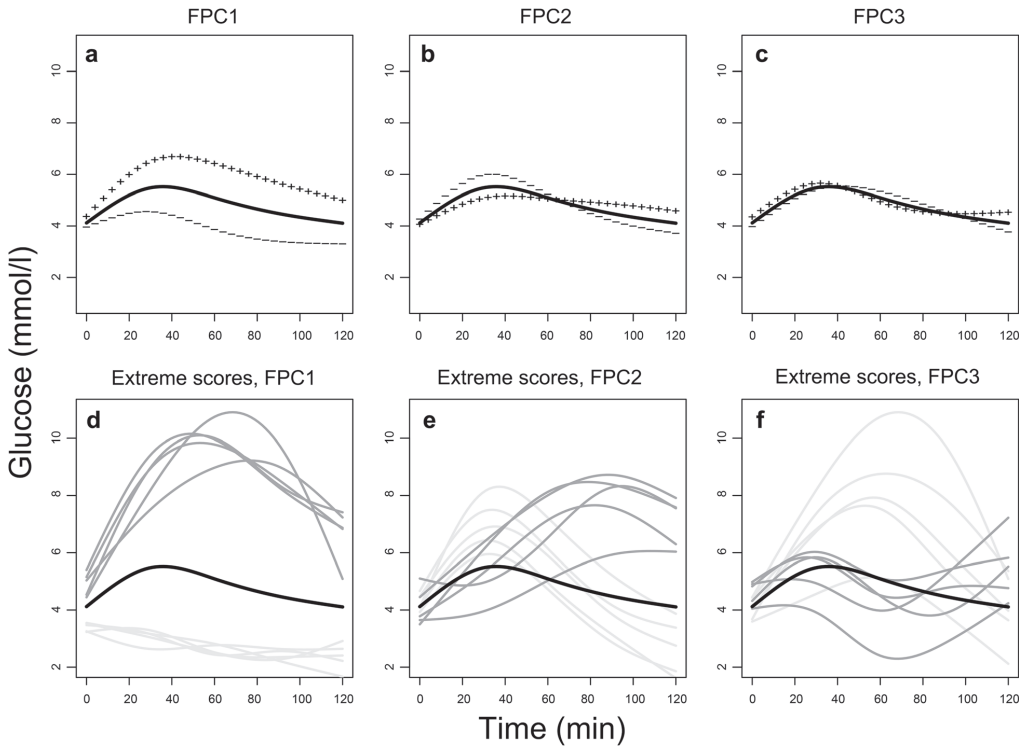


Figure 2 Results from the FPCA. a-c shows the mean of the fitted curves (solid line) and how the shape of an individual curve differs from the mean curve if a multiplum of the principal component curve (not shown) is added to (+ +) or subtracted from (– –) the mean curve. The multiplums correspond to one SD of the FPC1, FPC2 and FPC3 scores, respectively. **d-f** shows the mean of the fitted curves (black), and the individual curves for the five women with the highest positive scores (dark grey) and the five with the lowest negative scores (light grey) for each of the three FPCs.

"time to peak") and the oscillations in OGTT glucose curves (FPC3 "oscillations"), respectively. Women with high FPC1 scores had generally high glucose levels compared with the mean glucose level (Figure 2a). Women with high FPC2 scores had a longer than average time to peak, and it took longer for their glucose levels to return to normal postprandial levels (Figure 2b). Women with high FPC3 scores had curves that oscillated faster than the mean (Figure 2c). The plots of the five women with the highest and lowest scores for each of the FPCs (Figure 2d-f) highlighted these physiological interpretations. In sum, more than 99% of the total variation between the individual curves was explained by the first three FPCs, and further analyses were therefore restricted to these three FPCs.

For the majority of the women (89%), the entire OGTT glucose curve was between 2.5 and 7.8 mmol/l, while 6% had hypoglycaemic levels (values <2.5 mmol/l [32]) and three women were diagnosed with GDM. The 974 individual, fitted curves are grouped according to the lower and upper quartiles of the FPC1 and FPC2 scores in Figure 3. Women with high scores for both FPC1 and FPC2 had the highest glucose levels (Figure 3c), and these included the three women with GDM. Several women had OGTT glucose curve trajectories similar to those of the three GDM cases, but their curves descended below the GDM diagnosis threshold just before 2 h (Figure 3c).

Functional principal component scores vs simple summary measures

The FPCA transformed the five correlated OGTT measurements ($0.40 \leq r \leq 0.84$) into three uncorrelated FPC scores reflecting three distinct temporal features (Table 2). In contrast to fasting value, the 2-h value was positively associated with all three FPC scores ($0.37 \leq r \leq 0.79$). AUC was highly correlated with the FPC1 scores (r=0.999) but not with the FPC2 and FPC3 scores (r=−0.01 and r=0.05, respectively). The shape index was calculated as the 2-h value minus the 90-min value for 587 (60%) women, and as the 90-min value minus the 60-min value for 124 (13%) women. A total of 263 (27%) curves failed to meet the classification criteria of the shape index and were left out of these analyses. The shape index was most strongly associated with the FPC3 score (r=0.67). Pairwise scatter plots of these bivariate associations (not shown) showed that the three women classified as having GDM did not exhibit unusual FPC scores. Their FPC1 and FPC2 scores were high, but 33 other women had FPC1 scores in the same range, and 12 of them also had FPC2 scores above the upper quartile.

Functional analysis of variance

The means of the fitted curves differed between the four BMI categories (Figure 4a). While the curvature was similar, there were clear vertical shifts between the mean curves for normal weight, overweight and obese women. The functional CIs for the differences between underweight, overweight and obese women, as compared to normal weight women, are shown in Figure 4b. Pairwise comparisons of BMI categories showed the time periods of OGTT where the mean curves differed, as illustrated by the p curves in Figure 5. We found overall statistically significant differences between obese and overweight women (p<0.001), obese and normal weight women (p<0.001) and overweight and normal weight women (p<0.001). No statistically significant difference was found between underweight and normal weight women (p=0.26).

FANOVA vs ANOVA of simple summary measures

The results from ordinary ANOVA comparing the BMI categories in regard to fasting value, 2-h value or AUC were similar to those of the FANOVA comparisons. However, the shape index was only significantly different between obese and normal weight women (data not shown).

Multinomial regression with FPC scores

The means of the fitted curves at gestational weeks 14–16 for the seven pre-defined categories of 2-h values at gestational weeks 30–32 are shown in Figure 6. The women in the two upper categories (n=51) were all diagnosed with GDM at gestational weeks 30–32, but the mean curves in these two subgroups displayed different pathophysiology at gestational weeks 14–16. All women in the five lowest categories had a 2-h value below 7.8 mmol/l at gestational weeks 30–32, and were thus not diagnosed with GDM, but there were clear vertical shifts between their mean OGTT glucose curves at gestational weeks 14–16.

The results of the multinomial logistic regression analyses are shown in Table 3. The FPC1 scores and the AUC (Models 1 and 4, respectively) yielded nearly identical results, thus the results for AUC are not shown. We found that the mean FPC1 scores (and AUC) in the reference category were significantly different from the mean FPC1 scores in all other categories (all p<0.001), but that the mean FPC1 scores in subgroups of women with GDM were not significantly different (p=0.40). Also, the mean FPC1 scores in the lowest GDM category were not significantly different from the mean FPC1 scores in the closest non-GDM category (p=0.59). Similarly, no significant differences were found for fasting value, 2-h value or shape index in the three upper categories, i.e. between subgroups of women with and without GDM. In contrast, FPC2 scores discriminated between women who did and did not develop GDM, and between subgroups of women diagnosed with GDM later in pregnancy. The means of the FPC2 scores were significantly different between the three upper categories, p=0.01 and p=0.02, respectively. We also

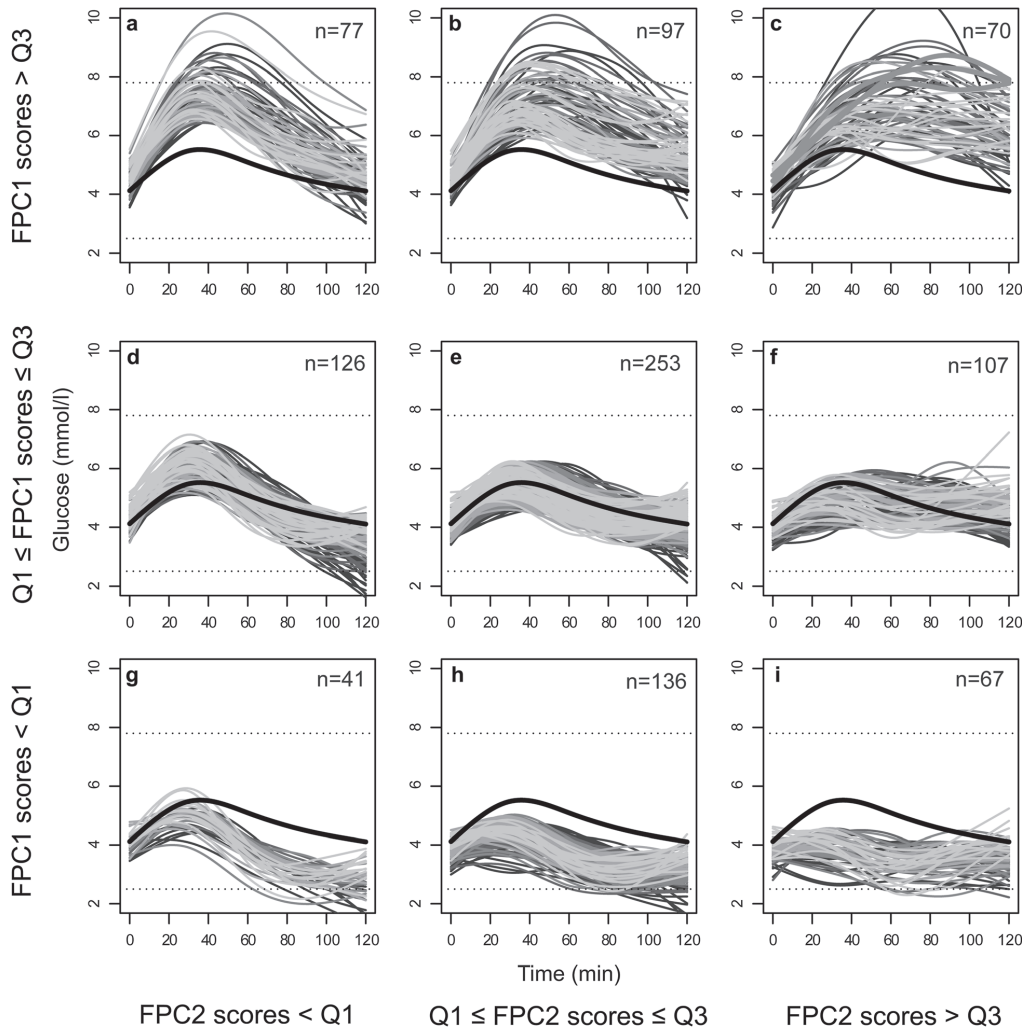


Figure 3 Individual curves. The figure shows the 974 individual, fitted curves classified according to the lower (Q1) and upper (Q3) quartiles of the FPC1 and FPC2 scores. The bold black curve is the overall mean of the fitted curves. Higher panels indicate higher FPC1 scores, and panels to the right represent higher FPC2 scores. The magnitudes of the FPC3 scores are represented using shades of grey: the lighter shades indicate higher FPC3 scores. The lower dashed line is 2.5 mmol/l, one possible cut-off for hypoglycaemia [32], and the upper dashed line is the diagnostic threshold for gestational diabetes, i.e. a 2-h value of 7.8 mmol/l [1]. The three women diagnosed with gestational diabetes are outlined with bold, grey lines in Figure 3c.

found a difference in the FPC3 scores between the two GDM categories (p=0.05) (Table 3).

Discussion

The present study demonstrated how information inherent in the shape of OGTT glucose curves can be extracted. The FDA approach yielded quantifiable shape entities with physiologically interpretable information that was not contained in the traditional simple summary measures. The extracted shape information differed significantly between women who did and did not develop GDM, and between subgroups of women diagnosed with GDM later in pregnancy, while various simple summary measures did not.

Table 2 Pearson correlation coefficients for OGTT measurements, FPC scores and simple summary measures (*n*=974)

OGTT	OGTT					FPC scores		
	Fasting	30 min	60 min	90 min	2 h	FPC1: "General level"	FPC2: "Time to peak"	FPC3: "Oscillation"
Fasting	1.00	0.44	0.40	0.41	0.42	0.47	−0.12	0.42
30 min		1.00	0.77	0.66	0.55	0.85	−0.47	0.19
60 min			1.00	0.84	0.70	0.96	−0.04	−0.22
90 min				1.00	0.80	0.93	0.31	−0.01
2 h					1.00	0.79	0.40	0.37
AUC	0.50	0.86	0.95	0.92	0.81	0.999	−0.01	0.05
Simple shape index[a]	−0.10	−0.34	−0.49	−0.41	0.12	−0.42	0.21	0.67

[a] *n*=711. Calculated as 2-h value minus 90-min value, or 90-min minus 60-min value [12].

The challenge of extracting shape information from glucose curves has been addressed by others [11-14], but these studies have focused on either simple shape indices or advanced parametric modelling. The present study is the first to use statistical tools and corresponding available software developed specifically for curves, to analyse OGTT data.

Our results were based on a large and relatively homogenous sample of healthy, pregnant women, but on a small number of glucose measurements per woman, as compared to those of an intravenous glucose tolerance test. One might expect to find even more physiologically interesting details and discriminating features of OGTT glucose curves, e.g. a larger number of FPCs with a substantial percentage of explained variability and more temporal details in the FPCs, in a more heterogeneous population with a more frequent OGTT sampling. For instance, our fitted curves could not reveal more than two peaks, but curves based on more densely sampled measurements over a longer time period than 2 h would likely show decreasingly oscillating curves rather than purely biphasic trajectories [14]. We therefore proposed the term "oscillating" as a qualitative description of OGTT glucose curves with more than one peak rather than using the term "biphasic", which has been used by others [12,14]. Furthermore, the classification of OGTT glucose curves as "biphasic", "monophasic" or "unclassified", involves several ad hoc conditions [12]. In the present study, we used FPC scores as continuous variables, as per general statistical recommendations, as this is the first choice of analysis in order to retain information and statistical power [33].

The mean of the fitted curves obtained from FDA (Figures 1, 2, 3) corresponded well with the familiar general shape of OGTT glucose curves [6,34,35]. In the literature in general, figures and analyses are usually based on the means at selected time points, with variability quantified by the SD or SE at the same time points, e.g. when comparing glucose responses [6]. In general, as seen in Figures 1, 2 and 3, the temporal mean undercommunicates the temporal variability. Although individual glucose curves have been presented in several publications [14,35,36], the variability in curve trajectories is highly under-reported, and thus largely unknown. As a result, the information indicated by the shape of OGTT glucose curves is rarely used in clinical practice, and only occasionally in research, although the standard

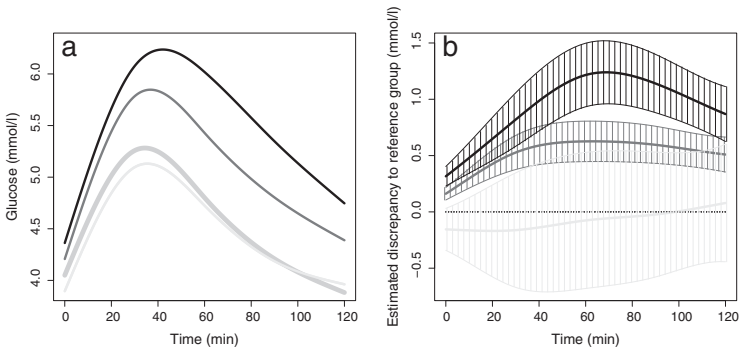


Figure 4 Results of the FANOVA. a shows the means of the fitted glucose curves for the BMI categories underweight (n=17, light grey curve), normal weight (n=588, bold grey curve), overweight (n=274, dark grey curve) and obese (n=87, black curve). **b** shows the estimated functional regression coefficients with corresponding CIs (shaded) and with normal weight as the reference category.

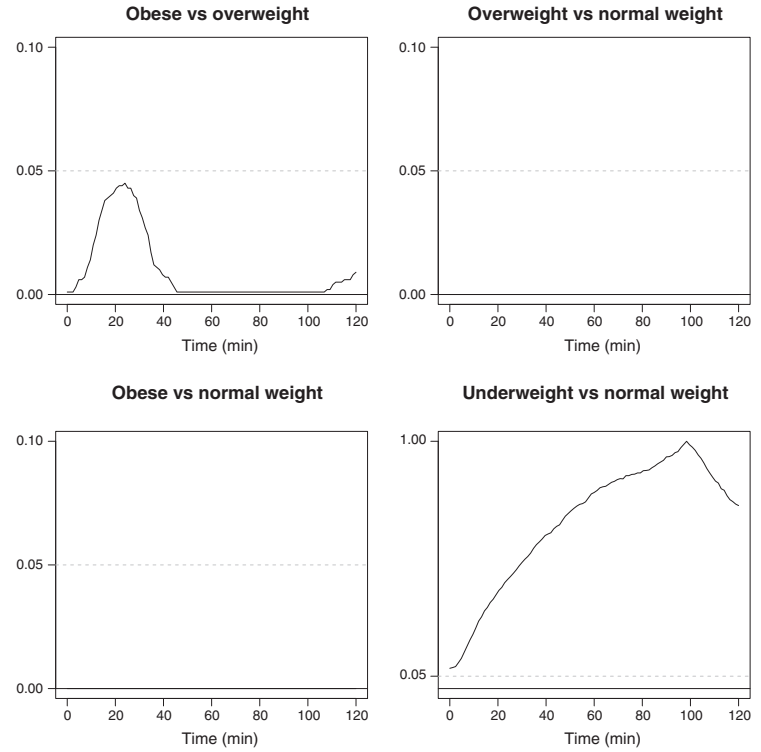


Figure 5 *p* **curves for pairwise comparisons of BMI categories using functional permutation tests.** The dashed line is the significance level of 0.05.

practice of taking repeated blood samples during OGTT suggests a focus on the curve. We have presented the individual, fitted curves in order to emphasise the heterogeneity between our study women and to provide a reference for OGTT glucose curves in healthy, pregnant women.

While a FPCA will decompose the variation between individual curves into a set of uncorrelated, temporal features, the clinical usefulness of this analysis depends on how the FPCs are interpreted. In this study, current insight into metabolism supported the interpretations of the FPCs as plausible and important physiological features. FPC1, which represented the general level and was the most important temporal feature of the curves, was almost perfectly correlated with AUC, and was significantly higher in women with high BMI. The fasting value and the 2-h value were also correlated with FPC1, but not as strongly as AUC. This is to be expected as a single measurement from a temporal phenomenon rarely describes the most essential temporal feature of the corresponding curve satisfactorily. Moreover, AUC is much better than the widely used fasting, or 2-h value in capturing the essential temporal information of OGTT glucose curves, which is consistent with results from previous studies [37-39]. The strongest association between the shape index and the FPC scores was found for FPC3 scores, which explained the smallest proportion of the total variance. This proportion was so small that FPC3 could have been left out of the analyses. We chose to include FPC3 for the comparison of FDA with the shape index. The shape index is based on an a priori classification of curves, involving an ad hoc set threshold for change. Many curves (27%) failed to meet the classification criteria and were left out of the analyses, resulting in a severe reduction of power and a biased representation of metabolic profiles in the study sample. Another, recently suggested shape index [13] is based on a rough approximation of the mean of the second order derivatives in the intervals between the measurements during the OGTT, giving a rough approximation of the total curvature. In the present study, FPC3 scores, representing the smallest proportion of the variance, quantified the amount of curvature. The shape feature of FPC3 was however less clear than for the first two

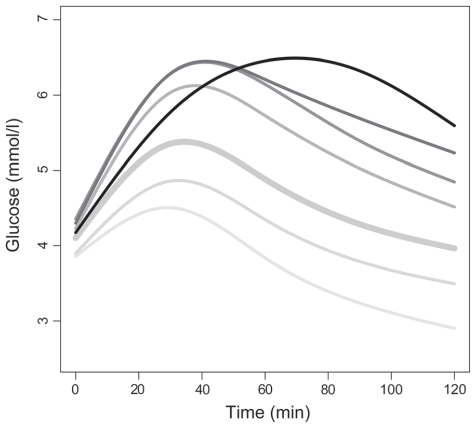


Figure 6 Means of glucose curves in first trimester, for different glucose categories later in pregnancy. The figure shows the means of the fitted glucose curves at gestational weeks 14–16, for different categories of 2-h values at gestational weeks 30–32. Darker lines indicate higher 2-h values. The 2-h glucose categories are <3.27, [3.27, 3.89), [3.89, 6.39), [6.39, 6.90), [6.90, 7.8), [7.8, 8.84) and ≥8.84 mmol/l.

components, and although it is possible that the third component might explain a larger part of the total variation if the sampling was more frequent and over a longer time period, this component should be used and interpreted with caution.

Glucose tolerance early in pregnancy has been found to predict glucose tolerance later in pregnancy [40]. The FPC1 scores, 2-h values and AUC differed significantly between groups of women without a GDM diagnosis at gestational weeks 30–32. However, only FPC2 scores were significantly different between women with and without GDM and only FPC2 and FPC3 scores differed significantly between diabetic women with the highest and second highest 2-h values in the third trimester. Thus, FPC1 or AUC alone did not capture all of the essential information about the differences in glucose metabolism. To distinguish curve trajectories reflecting deviating glucose tolerance from those considered normal, the information from FPC2 and FPC3 was necessary. A study of type 1 diabetes mellitus patients with islet transplantations showed that increased glucose AUC and time to peak C-peptide after metabolic testing were metabolic markers of islet allograft dysfunction [41], supporting the physiological importance of both FPC1 and FPC2 scores. The timing of the peak C-peptide was also found to be predictive of progression to type 1 diabetes mellitus in the Diabetes Prevention Trial [42].

The alternative to data-driven approaches such as FPCA for analysing full glucose curves is parametric modelling based on differential equation models of physiological mechanisms. Current concepts of blood glucose dynamics have been summarised in such models [14,43-45]. For instance, blood glucose levels and, hence, the shapes of glucose curves are affected by a number of key organs and physiologic processes that regulate the entry and removal of glucose from the blood [12,46]. A major disadvantage of parametric models is that estimating each person's individual parameters requires many measurements, often based on intravenous test procedures [47]. Although the use of OGTTs is debated [48], it is the simplest and most frequently used test procedure in larger studies because "gold-standard" intravenous procedures such as the euglycaemic clamp [49] are time-consuming, invasive and labour intensive.

Another important issue with parametric models of blood glucose regulation is the "closed loop" assumption, which can be hard to justify when modelling biological processes in the body because such processes are usually also susceptible to external influences. Diet, physical activity, obesity, changes in weight or visceral fat deposits, smoking and stress have all been shown to affect blood glucose levels [35] and external factors can have long-term effects on metabolism [50]. The genetic disposition of each individual adds to this complexity [51]. Finally, pregnancy causes alterations in a wide range of variables, including hormonal changes, insulin resistance and alterations in daily life habits. Nevertheless, parametric models seldom adjust for confounding by external variables [14,44,45]. Hence, even when parametric models seem to fit the data well, the error term for fit can include structural information not addressed in the predefined model, including information on the long-term effects of diet and the endocrine changes caused by pregnancy itself. This can make it difficult to validate the physiological theories underlying parametric models.

Although FDA or parametric modelling are the most natural approaches to glucose data for the study of glucose curves as single entities, there are alternatives to these analyses for the data presented in this article. For instance, the relation between BMI and glucose values could have been examined with a classical longitudinal data analysis with five repeated measurements per woman, with random effect of woman and modelling of the covariance structure. Also, instead of scores from FPCA, ordinary PCA scores based on the five glucose variables could be used as input to the regression analysis of glucose tolerance later in pregnancy. With only five measurements per curve, and measurements taken at the same time points for each woman, such traditional multivariate methods would be expected to extract similar information as the FDA. However, FDA is easier to apply in situations with more frequent sampling, sampling at unequal time points and missing data. In addition, FDA emphasizes the basic assumption about

Table 3 Results from four multinomial logistic regression analyses

2-h value, gestational weeks 30-32	*n*	Model 1: FPC1, FPC2 and FPC3 scores, gestational weeks 14–16*											
		FPC1 scores[a]				FPC2 scores				FPC3 scores			
		Mean (SD)	OR (95% CI)	*p*	*p*[b]	Mean (SD)	OR (95% CI)	*p*	*p*[b]	Mean (SD)	OR (95% CI)	*p*	*p*[b]
≥8.84	19	12.6 (13.5)	1.08 (1.04,1.13)	<0.001	0.40	5.7 (5.1)	1.36 (1.20,1.53)	<0.001	0.01	−0.7 (3.0)	0.87 (0.68,1.10)	0.23	0.05
[7.8,8.84)	32	11.1 (13.1)	1.11 (1.07,1.14)	<0.001	0.59	1.8 (4.7)	1.14 (1.04,1.25)	0.01	0.02	0.5 (1.6)	1.14 (0.95,1.37)	0.16	0.41
[6.90,7.8)	83	9.4 (12.8)	1.10 (1.07,1.12)	<0.001	0.02	−0.1 (3.1)	1.01 (0.95,1.08)	0.69	0.60	0.2 (1.9)	1.05 (0.93,1.19)	0.47	0.57
[6.39,6.90)	94	4.9 (9.6)	1.06 (1.04,1.09)	<0.001		−0.5 (3.8)	0.99 (0.93,1.06)	0.79		0.1 (1.8)	1.00 (0.89,1.13)	0.98	
[3.89,6.39)	601	−1.8 (9.7)	1	Ref		−0.2 (3.4)	1	Ref		0.1 (1.8)	1	Ref	
[3.27,3.89)	70	−6.9 (8.7)	0.94 (0.91,0.98)	<0.001	<0.01	0.0 (3.0)	0.98 (0.90,1.06)	0.62	0.07	−0.5 (1.7)	0.85 (0.74,0.98)	0.03	0.63
<3.27	23	−12.0 (9.8)	0.83 (0.78,0.90)	<0.001		−0.9 (3.8)	0.85 (0.73,0.98)	0.02		−0.8 (2.4)	0.80 (0.62,1.01)	0.07	
		Model 2: Fasting value gestational weeks 14-16*				Model 3: 2-h value gestational weeks 14-16*				Model 5: Simple shape index[c] gestational weeks 14-16*			
		Mean (SD)	OR (95% CI)	*p*	*p*[b]	Mean (SD)	OR (95% CI)	*p*	*p*[b]	Mean (SD)	OR (95% CI)	*p*	*p*[b]
≥8.84	19	4.1 (0.5)	2.00 (0.57,6.86)	0.28	0.55	5.5 (1.4)	3.40 (2.24,5.18)	<0.001	0.71	−0.80 (1.3)	0.53 (0.30,0.92)	0.03	0.36
[7.8,8.84)	32	4.1 (0.3)	3.17 (1.22,8.01)	0.02	0.54	5.3 (1.4)	3.11 (2.24,4.33)	<0.001	0.07	−0.47 (0.8)	0.73 (0.46,1.17)	0.19	0.92
[6.90,7.8)	83	4.2 (0.4)	4.43 (2.26,7.71)	<0.001	0.03	4.9 (1.1)	2.25 (1.79,2.84)	<0.001	0.01	−0.47 (0.8)	0.71 (0.52,0.97)	0.03	0.84
[6.39,6.90)	94	4.1 (0.4)	1.87 (0.99,3.25)	0.04		4.4 (0.9)	1.58 (1.26,1.98)	<0.001		−0.51 (0.7)	0.74 (0.55,1.00)	0.05	
[3.89,6.39)	601	4.0 (0.4)	1	Ref		4.0 (0.9)	1	Ref		−0.29 (0.7)	1	Ref	
[3.27,3.89)	70	3.8 (0.3)	0.32 (0.15,0.65)	<0.01	0.63	3.5 (0.8)	0.57 (0.42,0.78)	<0.001	<0.01	−0.28 (0.7)	0.95 (0.66,1.37)	0.80	0.18
<3.27	23	3.8 (0.4)	0.23 (0.09,0.93)	0.01		2.9 (0.4)	0.24 (0.14,0.41)	<0.001		−0.33 (1.1)	0.60 (0.34,1.07)	0.09	

[a] The FPC1 scores in model 1 and the AUC in model 4 yielded nearly identical results and the AUC results are thus not shown.
[b] *p* values from pairwise comparison between adjacent groups.
[c] *n*=711. Calculated as 2-h value minus 90-min value, or 90-min minus 60-min value [12].
* Categories of 2-h values in the third trimester is the response variable and OGTT characteristics in gestational weeks 14–16 are explanatory variables. All models are adjusted for BMI in gestational weeks 14–16.

continuity of the underlying process and its derivatives, and opens for analysis of the derivatives of the curves.

Contrary to general statistical advice [33], we have categorised two continuous variables in the analyses. An important aim of the present work was to introduce FDA and its benefits to a clinical audience. To ease the presentation of FDA, we chose to categorise BMI and the 2-h glucose at gestational weeks 30–32, based on the use of these variables in clinical practice. Different BMI categories are assumed to represent different risk groups [27], and BMI categories are frequently reported in clinical literature. The categorised BMI variable was therefore used in the analyses, although functional regression with BMI as a continuous variable would be preferable from a statistical point of view [33], especially as there were no obvious signs of nonlinearity (Figure 4a). The categorisation of the 2-h glucose value at gestational weeks 30–32, in contrast, revealed important non-linear relations (Figure 6). As an alternative to the multinomial logistic regression model, a regression model with the 2-h value as a continuous response variable could have been used.

The women in the cohort underwent two OGTTs, but only one was considered functional in the present work. We chose the 2-h value in third trimester as the main outcome instead of the entire curve in third trimester, due to the clinical relevance of this value in pregnancy care. As glucose curves are not commonly used, inference about the 2-h value would better illustrate the usefulness of information from FDA for a maternal pregnancy outcome in clinical practice.

Continuous glucose monitoring devices allow for more frequent glucose sampling over longer periods and might increasingly be used in future studies and in individual patient care to obtain OGTT measurements and measurements of glucose profiles in daily life. An increasing use of continuous glucose monitoring advocates the use of statistical tools that can properly analyse the continuous stream of data by providing curves that may be subjected to FDA as illustrated in the current work.

Furthermore, comparison of curve shape information from individuals with insulin resistance or beta cell failure might reveal whether curve features can distinguish between these two main processes that lead to the development of diabetes. Also, the curve shape information as obtained by FPCA in early pregnancy has the potential to predict complications in later pregnancy better than simple summary measures.

Our work shows that the FDA approach worked well, despite the very limited number of measurements for each participant. Dynamic, physiological processes will often be represented by scarcely sampled measurements, especially when repeated blood samples are required. In addition to glucose regulation, other examples where an FDA approach can be valuable include diurnal measurements of hormone regulation, metabolic changes during or after meals, or after physical exercise. The presented techniques should therefore also be explored in studies of metabolic disorders in non-pregnant populations.

Conclusions

In conclusion, the FDA approach was superior to traditional analyses of OGTT data in terms of providing physiologically interpretable and important temporal information, and in terms of differentiating between women who did and did not develop GDM during pregnancy. We recommend the FDA approach for the analysis of glucose data sampled repeatedly during glucose tolerance testing, or continuous glucose monitoring, to capitalise on important information that would otherwise be lost.

Appendix A

A.1. Curve fitting in functional data analysis

Let $y_i(t_j)$ be the measurement from individual i at time t_j, $i = 1, \ldots, n$ and $j = 1, \ldots, J$. In our OGTT data, $n = 974$ and $J = 5$. To each individual set of observations, $y_i(t_j)$, $j = 1, \ldots, J$, we fit a continuous, smooth function $x_i(t)$, spanning the observed time range. In our OGTT data, $t \in [0, 120]$. The estimation of the continuous curves $x_i(t)$ from data points $y_i(t_j)$ is based on the measurement model

$$y_i(t_j) = x_i(t_j) + \varepsilon_{ij}, \tag{1}$$

where $x_i(t_j)$ is x_i evaluated at time t_j and $\varepsilon_{ij} \sim N(0, \sigma^2)$ is an error term. It can be shown that a smooth curve is well approximated by a linear combination of a set of smooth basis functions $\phi_k(t)$, $k = 1, \ldots, K$,

$$x_i(t) \approx \sum_{k=1}^{K} c_{ki}\phi_k(t) = \mathbf{c}_i{}^T \phi(t), \tag{2}$$

where c_{ki} is the coefficient for the k^{th} basis function, $\mathbf{c}_i = (c_{1i}, \ldots, c_{Ki})$, and $\phi(t) = (\phi_1(t), \ldots, \phi_K(t))$. We apply B-spline basis functions, placing a knot at each of the J time points. With $\phi_k(t_j)$ denoting the k^{th} basis function evaluated at time t_j, substituting (2) into (1) yields

$$\begin{bmatrix} y_1(t_1) & \cdots & y_n(t_1) \\ \vdots & & \vdots \\ y_1(t_J) & \cdots & y_n(t_J) \end{bmatrix} = \begin{bmatrix} \phi_1(t_1) & \cdots & \phi_K(t_1) \\ \vdots & & \vdots \\ \phi_1(t_J) & \cdots & \phi_K(t_J) \end{bmatrix} \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{K1} & \cdots & c_{Kn} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1n} \\ \vdots & & \vdots \\ \varepsilon_{K1} & \cdots & \varepsilon_{Kn} \end{bmatrix}, \tag{3}$$

which in matrix notation reads

$$\mathbf{Y} = \mathbf{\Phi C} + \mathbf{E},$$

with $\mathbf{Y}$, $\mathbf{\Phi}$, $\mathbf{C}$ and $\mathbf{E}$ defined from (3). Here $\mathbf{Y}$ is the $J \times n$ matrix of observed blood glucose measurements; $\mathbf{\Phi}$ is the $J \times K$ matrix of the values of the K basis functions evaluated at times t_j, and $\mathbf{E}$ the $J \times n$ matrix of

error terms. Finally, $\mathbf{C}$ is the $K \times n$ matrix of unknown linear coefficients c_{ki}, which we estimate by minimising the penalised least squares expression

$$(\mathbf{Y}-\mathbf{\Phi C})^T(\mathbf{Y}-\mathbf{\Phi C}) + \lambda \mathbf{C}^T\mathbf{R}\mathbf{C}.$$

The penalty term, $\lambda \mathbf{C}^T\mathbf{R}\mathbf{C}$, where λ is a smoothing parameter that defines the degree of regularisation, is added to compensate for random error, and is based on the total curvature of the fitted curve,

$$\mathbf{R} = \int D^2\phi(s) D^2\phi^T(s) ds,$$

where $D^2\phi(s)$ is the second derivative of the vector of basis functions $\phi(t)$. The smoothing parameter $\lambda \in [0, \infty)$ is estimated by optimising a generalised cross-validation criterion. For more detail, see publications by Ramsay et al [17,18].

Appendix B

B.1. Functional principal component analysis

Functional principal component analysis (FPCA) can be viewed as rotating functional data to optimal empirical continuous basis functions, referred to as functional principal component (FPC) curves [17,18]. Associated with each FPC curve are individual FPC scores. These quantify how much the individual, fitted curves differ from the mean curve, in terms of the temporal pattern described by each FPC curve.

An FPC curve $\xi_\kappa(t)$ and its corresponding FPC scores $z_{\kappa i}$, $\kappa = 1, \ldots, K$, for individuals $i = 1, \ldots, n$, are estimated simultaneously by finding a weight function $\xi(t)$ defined over the same range of t as the functional data $x_i(t)$, maximising the variance of the corresponding individual FPC scores z_i, given by $z_i = \int \xi(t) x_i(t) dt$, subject to constraints. The first FPC, $\xi_1(t)$, is found by maximising the variance of the principal component scores z_{1i} subject to the constraint $\int \xi_1(t)^2 dt = 1$. Consecutive FPCs are defined similarly under the additional constraint of being orthogonal to the already extracted FPCs. For more detail, see publications by Ramsay et al [17,18].

Appendix C

C.1. Functional analysis of variance

Functional analysis of variance (FANOVA) is a method for studying the difference between the functional means of fitted curves in mutually exclusive subgroups of the study sample.

Consider a categorisation of the study sample into $g = 1, \ldots, G$ categories, e.g. BMI categories. Let L_g be the sample size in category g. We model the l^{th} OGTT glucose curve, $l = 1, \ldots, L_g$ in the g^{th} category, $x_{lg}(t)$, as

$$x_{lg}(t) = \beta_{\text{ref}}(t) + \beta_g(t) + \varepsilon_{lg}(t).$$

Here $\beta_{\text{ref}}(t)$ is the mean of the fitted curves in the reference category, $\beta_g(t)$ the difference between the mean curve in the g^{th} category and the reference category, and $\varepsilon_{ig}(t)$ the individual residual curve. The estimated group mean curve differences $\hat{\beta}_g(t), g = 1, \ldots, G,$ called the FANOVA coefficients, are based on the fitted curves described in Appendix A. They are also functions over the same t range.

Differences between categories can be evaluated by functional CIs for the FANOVA coefficients, corresponding $p(t)$ curves and overall p values from permutation F tests. The presented permutation tests are based on 1000 permutations of the fitted curves in two different categories. The CIs and $p(t)$ curves are calculated point-wise over the t range, using the estimated F-ratio $FR(t) = \frac{MRS(t)}{MSE(t)}$, calculated as the ratio of residual variance, $MRS(t)$, to predicted variance, $MSE(t)$. The permutation distribution is found for the point-wise F-statistic, giving CIs and $p(t)$ curves over the t range, and for the maximal value of the point-wise F-statistic, giving an overall p value. For more detail, see publications by Ramsay et al [17,18].

Additional file

Additional file 1: R script for functional data analysis of glucose curves.

Abbreviations

2-h: Two-hour; ANOVA: Analysis of variance; AUC: Area under the curve; BMI: Body mass index (kg/m^2); CI: Confidence interval; EDTA: Ethylenediaminetetraacetic acid; FANOVA: Functional analysis of variance; FDA: Functional data analysis; FPC: Functional principal component; FPCA: Functional principal component analysis; GDM: Gestational diabetes mellitus; OGTT: Oral glucose tolerance test; SD: Standard deviation.

Competing interests

The authors declare that they have no competing interests.

Authors' contribution

TH, JB, NV and KG contributed to the design of the STORK study and acquisition of data. KFF, JR and MBV designed the data analysis and KFF analysed the data. All authors contributed to the statistical or clinical interpretation of the results, writing and revising the manuscript, and approved the final version.

Acknowledgements

We gratefully acknowledge the help of each woman who participated and all the professional personnel at Oslo University Hospital Rikshospitalet who made this study possible. This study was supported by the Norwegian Health Association; the Faculty of Medicine, University of Oslo; and the Department of Obstetrics and Gynaecology and the National Resource Centre for Women's Health, Oslo University Hospital Rikshospitalet. Roche Diagnostics, Norway supplied glucometers and test strips free of charge.

Author details

[1]Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Boks 1122, Blindern 0317, Oslo, Norway. [2]Norwegian Resource Centre for Women's Health, Division of Obstetrics and Gynaecology, Oslo University Hospital, Rikshospitalet, Norway. [3]Section of Specialised Endocrinology, Department of Medicine, Oslo University Hospital, Rikshospitalet, Norway.

[4]Faculty of Clinical Medicine, University of Oslo, Rikshospitalet, Norway. [5]Division of Obstetrics and Gynaecology, Oslo University Hospital, Rikshospitalet, Norway.

Received: 13 September 2012 Accepted: 8 January 2013
Published: 17 January 2013

References

1. Alberti KGMM, Zimmet PZ: **Definition, diagnosis and classification of diabetes mellitus and its complications part 1: Diagnosis and classification of diabetes mellitus - Provisional report of a WHO consultation.** *Diabet Med* 1998, **15:**539–553.
2. World Health Organization: *Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia. Report of a WHO/IDF consultation*. Geneva: World Health Organization; 2006.
3. American Diabetes Association: **Standards of Medical Care in Diabetes-2011.** *Diabetes Care* 2011, **34**(Suppl 1):S11–S61.
4. HAPO Study Cooperative Research Group, Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, Hadden DR, McCance DR, Hod M, McIntyre HD, Oats JJ, Persson B, Rogers MS, Sacks DA: **Hyperglycemia and adverse pregnancy outcomes.** *N Engl J Med* 2008, **358:**1991–2002.
5. Levin RJ: **Carbohydrates**. In *Modern nutrition in health and disease*. 9th edition. Edited by Shils ME, Olson JA, Shine M, Ross AC. Baltimore: Lippincott Williams & Wilkins; 1999:49–65.
6. Jenkins DJA, Woelver TMS, Jenkins AL: **Fiber and other dietary factors affecting nutrient absorption and metabolism**. In *Modern nutrition in health and disease*. 9th edition. Edited by Shils ME, Olson JA, Shine M, Ross AC. Baltimore: Lippincott Williams & Wilkins; 1999:679–698.
7. MedlinePlus: *Glucose tolerance test*. http://www.nlm.nih.gov/medlineplus/ency/article/003466.htm.
8. Norwegian Directorate of Health: *National guidelines. Diabetes. Prevention, diagnostics and treatment. (In Norwegian.)*. http://helsedirektoratet.no/publikasjoner/nasjonale-faglige-retningslinjer—diabetes-brukerversjon/Publikasjoner/nasjonal-faglig-retningslinje—diabetes-brukerversjon.pdf.
9. Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, Pollak M, Regensteiner JG, Yee D: **Diabetes and cancer: a consensus report.** *Diabetes Care* 2010, **33:**1674–1685.
10. Pastor-Barriuso R, Guallar E, Coresh J: **Transition models for change-point estimation in logistic regression.** *Stat Med* 2003, **22:**1141–1162.
11. Zhou WB, Gu YY, Li H, Luo M: **Assessing 1-h plasma glucose and shape of the glucose curve during oral glucose tolerance test.** *Eur J Endocrinol* 2006, **155:**191–197.
12. Tschritter O, Fritsche A, Shirkavand F, Machicao F, Haring H, Stumvoll M: **Assessing the shape of the glucose curve during an oral glucose tolerance test.** *Diabetes Care* 2003, **26:**1026–1033. Erratum 27:1855.
13. Tura A, Morbiducci U, Sbrignadello S, Winhofer Y, Pacini G, Kautzky-Willer A: **Shape of glucose, insulin, C-peptide curves during a 3-h oral glucose tolerance test: any relationship with the degree of glucose tolerance?** *Am J Physiol Regul Integr Comp Physiol* 2011, **300:**R941–R948.
14. Trujillo-Arriaga HM, Roman-Ramos R: **Fitting and evaluating the glucose curve during a quasi continuous sampled oral glucose tolerance test.** *Comput Biol Med* 2008, **38:**185–195.
15. Siegmund T, Rad NT, Ritterath C, Siebert G, Henrich W, Buhling KJ: **Longitudinal changes in the continuous glucose profile measured by the CGMS in healthy pregnant women and determination of cut-off values.** *Eur J Obstet Gynecol Reprod Biol* 2008, **139:**46–52.
16. Yogev Y, Ben-Haroush A, Chen R, Rosenn B, Hod M, Langer O: **Diurnal glycemic profile in obese and normal weight nondiabetic pregnant women.** *Am J Obstet Gynecol* 2004, **191:**949–953.
17. Ramsay JO, Silverman BW: *Functional data analysis*. 2nd edition. New York: Springer; 2005.
18. Ramsay JO, Hooker G, Gray J: *Functional data analysis with R and MATLAB*. New York: Springer; 2009.
19. Ramsay JO: *Functional data analysis*. http://www.functionaldata.org.
20. Viviani R, Gron G, Spitzer M: **Functional principal component analysis of fMRI data.** *Hum Brain Mapp* 2005, **24:**109–129.
21. West RM, Harris K, Gilthorpe MS, Tolman C, Will EJ: **Functional data analysis applied to a randomized controlled clinical trial in hemodialysis patients describes the variability of patient responses in the control of renal anemia.** *J Am Soc Nephrol* 2007, **18:**2371–2376.
22. Coffey N, Harrison AJ, Donoghue OA, Hayes K: **Common functional principal components analysis: A new approach to analyzing human movement data.** *Hum Mov Sci* 2011, **30:**1144–1166.
23. Duhamel A, Devos P, Bourriez JL, Preda C, Defebvre L, Beuscart R: **Functional data analysis for gait curves study in Parkinson's disease.** *Stud Health Technol Inform* 2006, **124:**569–574.
24. Voldner N, Frøslie KF, Bo K, Haakstad L, Hoff C, Godang K, Bollerslev J, Henriksen T: **Modifiable determinants of fetal macrosomia: role of lifestyle-related factors.** *Acta Obstet Gynecol Scand* 2008, **87:**423–429.
25. Voldner N: ***Modifiable determinants of newborn macrosomia and birth complications.*** *PhD thesis. University of Oslo, Faculty of Medicine.*; 2010.
26. Frøslie KF, Godang K, Bollerslev J, Henriksen T, Røislien J, Veierød MB, Qvigstad E: **Correction of an unexpected increasing trend in glucose measurements during 7 years recruitment to a cohort study.** *Clin Biochem* 2011, **44:**1483–1486.
27. World Health Organization: **Obesity: Preventing and managing the global epidemic**. In *Report of a WHO Consultation*, WHO Technical Report Series, Volume 894. Geneva: World Health Organization; 2000.
28. International Association of Diabetes and Pregnancy Study Groups: **International Association of Diabetes and Pregnancy Study Groups Recommendations on the Diagnosis and Classification of Hyperglycemia in Pregnancy.** *Diabetes Care* 2010, **33:**676–682.
29. American Diabetes Association: **Diagnosis and Classification of Diabetes Mellitus.** *Diabetes Care* 2012, **35**(Suppl 1):S64–S71.
30. Hosmer DW, Lemeshow S: *Applied logistic regression*. 2nd edition. New York: Wiley; 2000.
31. *The R Foundation for Statistical Computing: R verison 2.13.0 (2011.04.13)*. http://www.r-project.org.
32. Hvidberg A, Fanelli CG, Hershey T, Terkamp C, Craft S, Cryer PE: **Impact of recent antecedent hypoglycemia on hypoglycemic cognitive dysfunction in nondiabetic humans.** *Diabetes* 1996, **45:**1030–1036.
33. Royston P, Altman DG, Sauerbrei W: **Dichotomizing continuous predictors in multiple regression: a bad idea.** *Stat Med* 2006, **25:**127–141.
34. Polonsky KS, Given BD, Hirsch LJ, Tillil H, Shapiro ET, Beebe C, Frank BH, Galloway JA, Van Cauter E: **Abnormal Patterns of Insulin-Secretion in Non-Insulin-Dependent Diabetes-Mellitus.** *N Eng J Med* 1988, **318:**1231–1239.
35. Freckmann G, Hagenlocher S, Baumstark A, Jendrike N, Gillen RC, Rössner K, Haug C: **Continuous glucose profiles in healthy subjects under everyday life conditions and after different meals.** *J Diabetes Sci Technol* 2007, **1:**695–703.
36. Kerssen A, de Valk HW, Visser GHA: **Day-to-day glucose variability during pregnancy in women with Type 1 diabetes mellitus: Glucose profiles measured with the Continuous Glucose Monitoring System.** *BJOG* 2004, **111:**919–924.
37. Sosenko JM, Palmer JP, Greenbaum CJ, Mahon J, Cowie C, Krischer JP, Chase HP, White NH, Buckingham B, Herold KC, Cuthbertson D, Skyler JS: **Diabetes Prevention Trial-Type 1 Study Group: Increasing the accuracy of oral glucose tolerance testing and extending its application to individuals with normal glucose tolerance for the prediction of type 1 diabetes - The Diabetes Prevention Trial-Type 1.** *Diabetes Care* 2007, **30:**38–42.
38. Ramachandran R, Gravenstein KS, Metter EJ, Egan JM, Ferrucci L, Chia CW: **Selective contribution of regional adiposity, skeletal muscle, and adipokines to glucose disposal in older adults.** *J Am Geriatr Soc* 2012, **60:**707–712.
39. Weijers RNM, Bekedam DJ, Goldschmidt HMJ, Smulders YM: **The clinical usefulness of glucose tolerance testing in gestational diabetes to predict early postpartum diabetes mellitus.** *Clin Chem Lab Med* 2006, **44:**99–104.
40. Phaloprakarn C, Tangjitgamol S: **Use of oral glucose tolerance test in early pregnancy to predict late-onset gestational diabetes mellitus in high-risk women.** *J Obstet Gynaecol Res* 2008, **34:**331–336.
41. Baidal DA, Faradji RN, Messinger S, Froud T, Monroy K, Ricordi C, Alejandro R: **Early metabolic markers of islet allograft dysfunction.** *Transplantation* 2009, **87:**689–697.
42. Sosenko JM, Palmer JP, Rafkin LE, Krischer JP, Cuthbertson D, Greenbaum CJ, Eisenbarth G, Skyler JS, Diabetes Prevention Trial-Type 1 Study Group: **Trends of earlier and later responses of C-peptide to oral glucose challenges with progression to type 1 diabetes in diabetes prevention trial-type 1 participants.** *Diabetes Care* 2010, **33:**620–625.
43. Steele R: **Influences of glucose loading and of injected insulin on hepatic glucose output.** *Ann N Y Acad Sci* 1959, **82:**420–430.

44. Turner RC, Holman RR, Matthews D, Hockaday TDR, Peto J: **Insulin deficiency and insulin resistance interaction in diabetes - estimation of their relative contribution by feedback analysis from basal plasma-insulin and glucose-concentrations.** *Metab Clin Experimental* 1979, **28:**1086–1096.
45. Bergman RN, Phillips LS, Cobelli C: **Physiologic evaluation of factors controlling glucose-tolerance in man - measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose.** *J Clin Invest* 1981, **68:**1456–1467.
46. Andersson EA, Pilgaard K, Pisinger C, Harder MN, Grarup N, Færch K, Sandholt C, Poulsen P, Witte DR, Jørgensen T, Vaag A, Pedersen O, Hansen T: **Do gene variants influencing adult adiposity affect birth weight? A population-based study of 24 loci in 4,744 Danish individuals.** *PLoS One* 2010, **5:**e14190.
47. Wallace TM, Matthews DR: **The assessment of insulin resistance in man.** *Diabet Med* 2002, **19:**527–534.
48. Davidson MB: **Counterpoint: The oral glucose tolerance test is superfluous.** *Diabetes Care* 2002, **25:**1883–1885.
49. Defronzo RA, Tobin JD, Andres R: **Glucose clamp technique - method for quantifying insulin-secretion and resistance.** *Am J Physiol* 1979, **237:**E214–E223.
50. Nazare JA, de Rougemont A, Normand S, Sauvinet V, Sothier M, Vinoy S, Désage M, Laville M: **Effect of postprandial modulation of glucose availability: short- and long-term analysis.** *Br J Nutr* 2010, **103:**1461–1470.
51. Faerch K, Borch-Johnsen K, Holst JJ, Vaag A: **Pathophysiology and aetiology of impaired fasting glycaemia and impaired glucose tolerance: does it matter for prevention and treatment of type 2 diabetes?** *Diabetologia* 2009, **52:**1714–1723.

doi:10.1186/1471-2288-13-6
Cite this article as: Frøslie *et al.*: Shape information from glucose curves: Functional data analysis compared with traditional summary measures. *BMC Medical Research Methodology* 2013 **13:**6.

Supplementary material, Article III

```
# R script for the functional data analysis (FDA) of early pregnancy OGTT data, as presented in the article
# Functional data analysis extracted physiologically important information from the shape of oral glucose tolerance test curves
# in a prospective cohort of pregnant women
# by KF Frøslie et al.

	# NB You might have to install the following R packages:
	# install.packages("foreign")
	# install.packages("fda")
	# install.packages("car")
	# install.packages("mlogit")
	# install.packages("Epi")


	# Read SPSS data, sort data and remove missing

library(foreign)
spssdata	<- as.data.frame(read.spss("M:/ArticleFDA/STORKdatabase.sav"))
spssdata	<- spssdata[order(spssdata$id),]
g		<- na.omit(cbind(spssdata$g0m,spssdata$g30m,spssdata$g60m,spssdata$g90m,spssdata$g2h))   # dim: 974,5

breaksuse	<- c(0,30,60,90,120)
ng		<- length(g[,1])

	# Plot of raw data

matplot(breaksuse,t(g[1:ng,]), type="l", lty=1, col="dark grey",lw=1,xlab="Time in minutes", ylab="Glucose in mmol/l")

	# FDA, first step: Curve fitting

library(fda)
mybasis	<- create.bspline.basis(rangeval=c(0,120),norder=4, breaks=c(0,30,60,90,120))		# Create B-spline basis (Appendix A)

	# Optimalisation of lambda, according to the generalised cross-validation criterion (Appendix A)

loglam	<- seq(-30,20,0.05)
# loglam	<- seq(0,20,0.05)		# Example of alternativ range for loglam in case of localt minimum (see below)
nlam	<- length(loglam)
dfsave	<- rep(NA,nlam)
gcvsave	<- rep(NA,nlam)

for (ilam in 1:nlam)
	{
	Lambda		<- 10^loglam[ilam]
	fdParobj	<- fdPar(mybasis,2,lambda)
	smoothlist	<- smooth.basis(c(0,30,60,90,120),t(g),fdParobj)
	dfsave[ilam]	<- smoothlist$df
	gcvsave[ilam]	<- sum(smoothlist$gcv)
	}
```

```
        # Optimal value of lambda:

lambdaopt      <- 10^loglam[gcvsave==min(gcvsave)]

        # Plot of gcv vs loglambda (ensure that the minimum is not a local one; if so, choose an alternative range for loglam, see above)
X11()
par(mfrow=c(2,3))
plot(loglam,gcvsave,type="l")
plot(loglam,gcvsave,type="l",ylim=c(0,100000))
plot(loglam,gcvsave,type="l",ylim=c(0,10000),xlim=c(-5,20))
plot(loglam,gcvsave,type="l",ylim=c(0,2000),xlim=c(-5,7))
plot(loglam,gcvsave,type="l",ylim=c(1200,1400),xlim=c(-5,-4))
abline(h=min(gcvsave),lty=2)
plot(loglam,gcvsave,type="l",ylim=c(1200,1400),xlim=c(2,4))
abline(h=min(gcvsave),lty=2)


        # Optimal smoothing, according to gcv criterion (Appendix A)

fdParobj.opt   <- fdPar(mybasis,2,lambdaopt)
g.smooth.opt   <- smooth.basis(c(0,30,60,90,120),t(g),fdParobj.opt)
plot(g.smooth.opt,lty=1,col="grey")

        # Plot of smoothed/fitted curves (Figure 1)

par(mfrow=c(1,2),mar=c(4,5,0,0)+0.2)
n=5
matplot(breaksuse,t(g[1:(n),]), type="b",pch= 20,lty=1, col=gray(0.85),lw=1,
        xlab="Time (min)", ylab="Glucose (mmol/l)", ylim=c(1.5,12),cex.lab=2,cex.axis=1.5,cex=2)
plot(g.smooth.opt$fd[1:n], lty=1, col=gray(0.5),lw=1,add=TRUE,
        xlab="Time (min)", ylab="", ylim=c(1.5,12),cex.lab=2,cex.axis=1.5)
plot(g.smooth.opt$fd[1:ng], lty=1, col=gray(0.5),lw=1,add=FALSE,
        xlab="Time (min)", ylab="",cex.lab=2,cex.axis=1.5)
lines(mean(g.smooth.opt$fd[1:ng]),col="black",lw=4)

        # Functional principal component analysis, FPCA (Appendix B)

g.pca          <- pca.fd(g.smooth.opt$fd,nharm=3)   # We keep only the first 3 FPCs
par(mfrow=c(1,3))
plot.pca.fd(g.pca)                                  # Quick FPCA plot
plot(g.pca$harmonics)                               # NB: 'harmonics' is another word for 'FPCA curves'

        # Bivariate correlations in Table 2

gdata                 <- cbind(g.pca$scores,g,(0.5*g[,1]+g[,2]+g[,3]+g[,4]+0.5*g[,5]),(g[,5]-g[,4]))
colnames(gdata)       <- c("fpc1","fpc2","fpc3","g0m","g30m","g60m","g90m","g2h","auc","shape")
head(gdata)
round(cor(gdata,use="pairwise.complete.obs"),2)
```

```
       # Nice FPCA plot (Figure 2 a-c)

eval.pca       <- seq(0,120,5)                              # The vector eval.pca makes sure that the +'s and -'s are not too close
g.pca.mean     <- eval.fd(eval.pca, g.pca$meanfd)           # Vector of mean values
g.pca.points   <- eval.fd(eval.pca, g.pca$harmonics)        # Vectors of harmonics-values (no. of vectors equals the no. of harmonics)
g.pca.values   <- g.pca$values                              # The variances of the score variables. Defines the 'multiplum' in Figure 2

ylim.g  <- c(3,8.2)
par(mfcol=c(1,3), mar=c(2,4,1,1))
for(ii in 1:3){
plot(eval.pca,g.pca.mean, ylim=ylim.g, type="l",ylab=paste("Principal component curve",ii))
points(eval.pca,g.pca.mean + sqrt(g.pca.values[ii])*g.pca.points[,ii], pch="+",col="dark grey")
points(eval.pca,g.pca.mean - sqrt(g.pca.values[ii])*g.pca.points[,ii], pch="-",col="dark grey")
}

round(g.pca.values/sum(g.pca.values),3)                     # Percentage explained variance

       # Functional analysis of variance, FANOVA (Appendix C)
       # Data preparation (remove missing)

library(car)
dim(spssdata)
data           <-spssdata[!is.na(spssdata$g0m)&
                          !is.na(spssdata$g30m)&
                          !is.na(spssdata$g60m)&
                          !is.na(spssdata$g90m)&
                          !is.na(spssdata$g2h)&
                          !is.na(spssdata$bmi),]

g.bmi          <- cbind(data$g0m,data$g30m,data$g60m,data$g90m,data$g2h)
bmi.cut        <- cut(data$bmi,breaks=c(0,18.5,25,30,100), labels=FALSE, include.lowest=TRUE, right=FALSE)     # categorise into WHO categories
bmi.cut        <- as.factor(na.omit(bmi.cut))
bmi.cut        <- recode(bmi.cut,"1='6'")                   # Make group 2 the reference category and code underweight as 6
table(bmi.cut)

bmi.smooth     <- smooth.basis(breaksuse,t(g.bmi),fdParobj.opt)
curves.bmi     <- bmi.smooth$fd

       # Mean glucose curves in the BMI categories (Fig. 4 a)

par(mfrow=c(1,2))
plot(mean.fd(curves.bmi[bmi.cut==6]),lw=4,ylim=c(3,8), col=gray(0.9),ylab="Glucose (mmol/l)",xlab="Time (min)",cex.lab=1.5)
plot(mean.fd(curves.bmi[bmi.cut==2]),lw=8,ylim=c(3.9,6.3), col=gray(0.8),add=TRUE )
plot(mean.fd(curves.bmi[bmi.cut==6]),lw=4,ylim=c(3.9,6.3), col=gray(0.9),add=TRUE)       # Plot underweight again to emphasise the crossing curves
plot(mean.fd(curves.bmi[bmi.cut==3]),lw=4,ylim=c(3.9,6.3), col=gray(0.5),add=TRUE )
plot(mean.fd(curves.bmi[bmi.cut==4]),lw=4,ylim=c(3.9,6.3), col=gray(0),add=TRUE )
```

```
m1.fanova             <- fRegress(curves.bmi ~ bmi.cut)           # The FANOVA
summary(m1.fanova$betaestlist)
X11()
par(mfrow=c(3,4))
lapply(m1.fanova$betaestlist, plot)

       # Plot of estimated beta curves only; a separate plot for the reference curve and an extra for the discrepancies (cf Fig. 4 b)

plot(m1.fanova$betaestlist$const,main="const")
plot(m1.fanova$betaestlist$bmi.cut.3,main="bmi 25-30",ylim=c(-1,1.5),col="red")
plot(m1.fanova$betaestlist$bmi.cut.4,main="bmi over 30",col="magenta",add=TRUE)
plot(m1.fanova$betaestlist$bmi.cut.6,main="bmi under 18.5",col="cyan",add=TRUE)

       # Plot of estimated beta curves with functional confidence intervals (editing not possible)

bmi.smooth.y2cMap     <- bmi.smooth$y2cMap
m1.fanova.yhat        <- m1.fanova$yhatfdobj$fd
m1.Errmat             <- t(g.bmi)-eval.fd(breaksuse,m1.fanova.yhat)
m1.SigmaE2            <- cov(t(m1.Errmat))
m1.fanova.std         <- fRegress.stderr(m1.fanova,bmi.smooth.y2cMap,m1.SigmaE2)

plotbeta(m1.fanova$betaestlist,m1.fanova.std$betastderrlist)

       # To make a plot like the one in Fig. 4 b, you need to make a modified version of the function plotbeta(); e.g. myplotbeta1()
       # Save it as an R script: myplotbeta1.R, and load it by

source("M:/ArticleFDA/myplotbeta1.R")
myplotbeta1(m1.fanova$betaestlist,m1.fanova.std$betastderrlist)    # (Figure 4 b)

       # Plots of dicrepancies between fitted curves and curves estimated from the model, as well as residuals

par(mfrow=c(2,2))
plot(curves.bmi,col=as.numeric(bmi.cut)+2,lty=1)
plot(m1.fanova$yhat$fd, lw=3,col="black",lty=1,add=TRUE)
yhatfd                <- m1.fanova$yhat$fd
plot(curves.bmi-yhatfd,col=as.numeric(bmi.cut)+2,lty=1)
res.matrix            <- eval.fd(breaksuse,curves.bmi)-eval.fd(breaksuse,yhatfd)
boxplot(t(res.matrix))

       # Permutation F tests for pairwise comparisons of BMI groups (Give the plots in Fig. 5)
       # Underweight vs normal weight: group 6 (bmi under 18.5) vs group 2 (Normal, bmi 18.5-25)

bmi.26         <- bmi.cut[(bmi.cut==2|bmi.cut==6)]
curves.bmi26   <- curves.bmi[(bmi.cut==2|bmi.cut==6)]
table(bmi.26)

cbasis         <- create.constant.basis(c(0,120))   # Make a constant basis
bmi.26         <- as.numeric(bmi.26)-1              # Recode factors "2" and "6" to the numbers 1 and 4; subtract 1 and get 0 and 3
bmi.26[bmi.26==3] <-1                               # Recode 3 to 1, i.e. 0 is the reference category and 1 is the category of interest
```

```
constfd        <- fd( matrix(1,1,length(bmi.26)),cbasis)
bmi26fd        <- fd( matrix(bmi.26,1,length(bmi.26)),cbasis)
xfdlist        <- list(constfd,bmi26fd )
betalist       <- list(fdParobj.opt,fdParobj.opt)

Fres12         <- Fperm.fd(curves.bmi26,xfdlist,betalist,nperm=1000)      # nperm should be large enough (here: 1000) to visualise the details
                                                                          # NB: Fperm.fd gives automatically a Permutation F Test plot
Fres12$pval
plot(Fres12$argvals,Fres12$pvals.pts, type="l")

       # Overweight vs normal: group 3 (bmi 25-30) vs group 2 (bmi 18.5-25)

bmi.23         <- bmi.cut[(bmi.cut==2|bmi.cut==3)]
curves.bmi23   <- curves.bmi[(bmi.cut==2|bmi.cut==3)]
table(bmi.23)
bmi.23         <- as.numeric(bmi.23)-1

constfd        <- fd( matrix(1,1,length(bmi.23)),cbasis)
bmi23fd        <- fd( matrix(bmi.23,1,length(bmi.23)),cbasis)
xfdlist        <- list(constfd,bmi23fd )
betalist       <- list(fdParobj.opt,fdParobj.opt)

Fres23         <- Fperm.fd(curves.bmi23,xfdlist,betalist,nperm=1000)
Fres23$pval
plot(Fres23$argvals,Fres23$pvals.pts, type="l")

       # Obese vs normal: group 4 vs group 2

bmi.24         <- bmi.cut[(bmi.cut==2|bmi.cut==4)]
curves.bmi24   <- curves.bmi[(bmi.cut==2|bmi.cut==4)]
table(bmi.24)
bmi.24 <- as.numeric(bmi.24)-1
bmi.24[bmi.24==2] <-1

constfd        <- fd( matrix(1,1,length(bmi.24)),cbasis)
bmi24fd        <- fd( matrix(bmi.24,1,length(bmi.24)),cbasis)
xfdlist        <- list(constfd,bmi24fd )
betalist       <- list(fdParobj.opt,fdParobj.opt)

Fres24         <- Fperm.fd(curves.bmi24,xfdlist,betalist,nperm=1000)
Fres24$pval
plot(Fres24$argvals,Fres24$pvals.pts, type="l")

       # Overweight vs obese: group 3 vs group 4

bmi.34         <- bmi.cut[(bmi.cut==3|bmi.cut==4)]
curves.bmi34   <- curves.bmi[(bmi.cut==3|bmi.cut==4)]
table(bmi.34)
bmi.34 <- as.numeric(bmi.34)-2
```

```
constfd        <- fd( matrix(1,1,length(bmi.34)),cbasis)
bmi34fd        <- fd( matrix(bmi.34,1,length(bmi.34)),cbasis)
xfdlist        <- list(constfd,bmi34fd )
betalist       <- list(fdParobj.opt,fdParobj.opt)

Fres34         <- Fperm.fd(curves.bmi34,xfdlist,betalist,nperm=1000)
Fres34$pval
plot(Fres34$argvals,Fres34$pvals.pts, type="l")

       # Analysing the effect of first trimester glucose curves on the 2-h value at wks 30-32 by multinomial regression
       # Data preparation (remove missing)

data   <- spssdata[   !is.na(spssdata$g0m)&
                      !is.na(spssdata$g30m)&
                      !is.na(spssdata$g60m)&
                      !is.na(spssdata$g90m)&
                      !is.na(spssdata$g2h)&
                      !is.na(spssdata$bmi)&
                      !is.na(spssdata$g2hwk30),]

       # Categorise into 7 categories

g.g2hwk30              <- cbind(data$g0m,data$g30m,data$g60m,data$g90m,data$g2h,data$bmi, data$g2hwk30)
g2hwk30.cut            <- cut(data$g2hwk30,breaks=c(0,3.27,3.89,6.39,6.90,7.8,8.84,100), labels=FALSE, include.lowest=TRUE, right=FALSE)
table(g2hwk30.cut)

g.g2hwk30.smooth       <- smooth.basis(breaksuse,t(g.g2hwk30[,-c(6,7)]),fdParobj.opt)
curves.g.g2hwk30       <- g.g2hwk30.smooth$fd

       # Mean glucose curves for groups of women in different glucose categories at gestational weeks 30-32 (Fig. 6)

plot(mean.fd(g.g2hwk30.smooth$fd[g2hwk30.cut==1]),  lw=4,ylim=c(2.5,7), col=gray(0.9),ylab="Glucose (mmol/l)",xlab="Time (min)",cex.lab=1.5)
plot(mean.fd(g.g2hwk30.smooth$fd[g2hwk30.cut==2]),  lw=4,ylim=c(3.9,6.4), col=gray(0.85),add=TRUE )
plot(mean.fd(g.g2hwk30.smooth$fd[g2hwk30.cut==3]),  lw=8,ylim=c(3.9,6.4), col=gray(0.8),add=TRUE )
plot(mean.fd(g.g2hwk30.smooth$fd[g2hwk30.cut==4]),  lw=4,ylim=c(3.9,6.4), col=gray(0.7),add=TRUE )
plot(mean.fd(g.g2hwk30.smooth$fd[g2hwk30.cut==5]),  lw=4,ylim=c(3.9,6.4), col=gray(0.55),add=TRUE )
plot(mean.fd(g.g2hwk30.smooth$fd[g2hwk30.cut==6]),  lw=4,ylim=c(3.9,6.4), col=gray(0.4),add=TRUE )
plot(mean.fd(g.g2hwk30.smooth$fd[g2hwk30.cut==7]),  lw=4,ylim=c(3.9,6.4), col=gray(0),add=TRUE )

       # FPCA of the reduced data set

g.g2hwk30.pca          <- pca.fd(g.g2hwk30.smooth$fd,nharm=3)
plot.pca.fd(g.g2hwk30.pca)
plot(g.g2hwk30.pca$harmonics)
g.g2hwk30.pca$values
round(g.g2hwk30.pca$values/sum(g.g2hwk30.pca$values),3)

mndata                 <- as.data.frame(na.omit(cbind(g.g2hwk30.pca$scores[,1:3],g2hwk30.cut,g.g2hwk30[,6])) )
colnames(mndata)       <- c("fpc1","fpc2","fpc3","g2hwk30kat","bmi")
mndata$g2hwk30kat      <- as.factor(mndata$g2hwk30kat)
```

```
       # Multinomial logistic regression

library(mlogit)
library(Epi)
mldata         <- mlogit.data(mndata, varying=NULL, choice="g2hwk30kat", shape="wide")
mref           <- mlogit(g2hwk30kat~1|fpc1+fpc2+fpc3+bmi, data = mldata, reflevel="3")
summary(mref)
exp(coef(mref))
mref           <- mlogit(g2hwk30kat~1|fpc1+fpc2+fpc3+bmi, data = mldata, reflevel="7")
summary(mref)
mref           <- mlogit(g2hwk30kat~1|fpc1+fpc2+fpc3+bmi, data = mldata, reflevel="6")
summary(mref)
mref           <- mlogit(g2hwk30kat~1|fpc1+fpc2+fpc3+bmi, data = mldata, reflevel="5")
summary(mref)
mref           <- mlogit(g2hwk30kat~1|fpc1+fpc2+fpc3+bmi, data = mldata, reflevel="4")
summary(mref)
mref           <- mlogit(g2hwk30kat~1|fpc1+fpc2+fpc3+bmi, data = mldata, reflevel="2")
summary(mref)

mref           <- mlogit(g2hwk30kat~1| g0m +bmi, data = mldata, reflevel="3")
summary(mref)
exp(coef(mref))
mref           <- mlogit(g2hwk30kat~1| g0m +bmi, data = mldata, reflevel="6")
summary(mref)
mref           <- mlogit(g2hwk30kat~1| g0m +bmi, data = mldata, reflevel="5")
summary(mref)
mref           <- mlogit(g2hwk30kat~1| g0m +bmi, data = mldata, reflevel="2")
summary(mref)

mref           <- mlogit(g2hwk30kat~1| g2h +bmi, data = mldata, reflevel="3")
summary(mref)
exp(coef(mref))
mref           <- mlogit(g2hwk30kat~1| g2h +bmi, data = mldata, reflevel="6")
summary(mref)
mref           <- mlogit(g2hwk30kat~1| g2h +bmi, data = mldata, reflevel="5")
summary(mref)
mref           <- mlogit(g2hwk30kat~1| g2h +bmi, data = mldata, reflevel="2")
summary(mref)

mref           <- mlogit(g2hwk30kat~1|shape+bmi, data = mldata, reflevel="3")
summary(mref)
exp(coef(mref))
mref           <- mlogit(g2hwk30kat~1|shape+bmi, data = mldata, reflevel="6")
summary(mref)
mref           <- mlogit(g2hwk30kat~1|shape+bmi, data = mldata, reflevel="5")
summary(mref)
mref           <- mlogit(g2hwk30kat~1|shape+bmi, data = mldata, reflevel="2")
summary(mref)
```

IV

Title:

Shape information in repeated glucose curves during pregnancy provided significant physiological information for neonatal outcomes.

Authors: Kathrine Frey Frøslie[*a, b], Jo Røislien[a], Elisabeth Qvigstad[c], Kristin Godang[c], Jens Bollerslev[c,d], Tore Henriksen[d,e], Marit B Veierød[a]

Affiliations:

[a] Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Norway

[b] Norwegian Resource Centre for Women's Health, Division of Obstetrics and Gynaecology, Oslo University Hospital Rikshospitalet, Norway

[c] Section of Specialised Endocrinology, Department of Medicine, Oslo University Hospital, Norway

[d] Faculty of Clinical Medicine, University of Oslo, Norway

[e] Division of Obstetrics and Gynaecology, Oslo University Hospital Rikshospitalet, Norway

*Corresponding author:

Kathrine Frey Frøslie

Department of Biostatistics, Institute of Basic Medical Sciences UiO, Boks 1122 Blindern, 0317 Oslo, Norway

Tel: +47 95 04 47 10 Fax: +47 22 85 13 13

Email address: k.f.froslie@medisin.uio.no

Abstract

Objective: To use multilevel functional principal component analysis to exploit the information inherent in the shape of longitudinally sampled glucose curves during pregnancy, and to analyse the impact of glucose curve characteristics on neonatal birth weight, percentage fat and cord blood C-peptide.

Study Design and Setting: A cohort study of healthy, pregnant women (n=884). They underwent two oral glucose tolerance tests (gestational weeks 14-16 and 30-32), which gave two glucose curves per woman.

Results: Glucose values were higher, and peaked later in third trimester than in early pregnancy. The curve characteristic "general glucose level" accounted for 91% of the variation across visits, and 72% within visits. The curve characteristics "timing of postprandial peak", and "oscillating glucose levels" accounted for a larger part of the variation within visits (15% and 8%), than across visits (7% and <2%). A late postprandial peak during pregnancy, and high general glucose levels in third trimester had significant, positive effects on birth weight ($p<0.05$). Generally high glucose levels during pregnancy had a significant, positive impact on neonatal percentage fat ($p=0.04$). High general glucose level in third trimester had a significant, positive impact on cord blood C-peptide ($p=0.004$).

Conclusion: Shape information in entire OGTT curves provides significant physiological information of importance for several outcomes, and may contribute to the understanding of the metabolic changes during pregnancy.

Introduction

High maternal glucose levels in pregnancy have adverse short-term and long-term health effects for both the mother and the child [1–5]. The Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) study investigated glucose intolerance less severe than that in overt diabetes mellitus, and demonstrated effects on the risk of adverse pregnancy outcomes [1]: Positive, linear effects were found for the fasting, one-hour (1-h) and two-hour (2-h) values from oral glucose tolerance tests (OGTTs). Other studies reporting associations between high maternal glucose levels and adverse pregnancy outcomes have used a variety of simple glucose measures, e.g. the fasting value, the 2-h value, impaired fasting glucose, gestational diabetes (GDM) diagnosis or HbA1c [6–12].

Changes in glucose metabolism during pregnancy include increasing insulin resistance and increasing gluconeogenesis in the liver [13]. Counterintuitive to this, longitudinal studies have reported a decrease in fasting glucose levels during pregnancy, particularly during the first trimester [13–15]. However, concomitantly with the decrease in fasting glucose, elevated postprandial levels during pregnancy have been reported [9,16,17]. Some studies have described glucose curves or glucose data at different gestational ages and longitudinal changes in these curves and data during pregnancy [9,16,18,19]. Few studies have analysed the impact of information in the shape of entire OGTT glucose curves [20], and except from one study [16],we are not aware of statistical analysis of longitudinal change in glucose curves during pregnancy. Also, few have studied the impact of change in glucose levels during pregnancy on neonatal outcomes [21,22].

Functional data analysis (FDA) is a collection of statistical methods developed to analyse curve data [23,24]. In FDA a set of temporal observations is treated as a single, functional object. The statistical analysis is based on this continuous function (curve), rather than on the original discrete data points. Information from the curve as a whole is extracted. We have demonstrated that glucose curves from OGTT at one time point in pregnancy (gestational weeks 14-16) extracted physiologically interpretable and clinically interesting characteristics of the glucose response that would otherwise be missed [20]. We now extend the analysis to study the shape inherent in glucose curves from two visits during pregnancy. To our knowledge, this is the first study to use all information in longitudinally collected glucose curves, and to analyse the effect of such information on neonatal outcomes.

The STORK study, a Norwegian prospective cohort study of 1031 healthy, pregnant women, provided OGTT data from gestational weeks 14-16 and 30-32 [25]. Using FDA methodology developed by Di et al [26] and Crainiceanu and Goldsmith [27], we performed a

multilevel FDA of the OGTT data, and extracted essential characteristics of the OGTT glucose curves from gestational weeks 14-16 and 30-32. We then studied the effect of these characteristics on the neonatal outcomes birth weight, percentage fat and C-peptide in cord blood.

Methods

Ethics Statement

The study was approved by the Regional Committee for Medical Research Ethics, Southern Norway, Oslo, Norway (reference number S-01191), and performed according to the Declaration of Helsinki. All participating women provided written informed consent.

Participants and data

The STORK study is a prospective cohort of 1031 healthy, Norwegian women of Scandinavian heritage who registered for obstetric care at Oslo University Hospital Rikshospitalet from 2001 to 2008. The overall aim of the study was to extend insights into maternal metabolic syndrome and determinants of foetal macrosomia [28]. Exclusion criteria were multiple pregnancies, known pre-gestational diabetes, and severe chronic diseases (pulmonary, cardiac, gastrointestinal or renal). Gestational age at inclusion was based on the Naegele's rule, and gestational ages at the other visits and at birth were based on routine ultrasound at weeks 17-19. Age, parity, smoking habits, height, weight, fasting insulin and a 75g OGTT were recorded at inclusion at weeks 14-16. Weight, fasting insulin and OGTT were also recorded at weeks 30-32.

Blood samples were drawn in the morning, between 0730 and 0830 after an overnight fast, and were obtained from veni-puncture in tubes containing Ethylenediaminetetraacetic acid (EDTA). Plasma glucose was measured immediately in a drop of fresh, whole EDTA blood. During the OGTT, blood samples were taken every 30 minute for 2 hours. Glucose measurements were done by the Accu-Chek Sensor (ACS) glucometer (Roche Diagnostics GmbH, Mannheim, Germany). Due to an unexpected increasing trend in the fasting glucose measurements over the 7 years of recruitment, all glucose measurements were de-trended prior to the analyses [29]. The umbilical cord blood was collected into EDTA tubes by the midwife, centrifuged for plasma separation and placed at -20 ºC for less than a month and stored long term at -80 ºC.

Women with premature births or non-complete OGTT data were excluded, giving a study sample of 884 women and their neonates (Figure 1).

Birth weight was recorded within two hours after the birth. In a subsample of the cohort, the percentage of neonatal body fat was measured by Dual-energy X-ray absorptiometry (DXA) scanning, and C-peptide in cord blood from the time of birth was measured (Figure 1) [30,31].

Data description

Descriptive statistics of registered data are presented as mean, standard deviation (SD) and range, or frequency and percentage (%).

Fitting continuous and individually smoothed curves

The OGTT measurements from the two visits were converted into 884 continuous, smooth OGTT glucose curves (hereafter called glucose curves), from gestational weeks 14-16, and 884 continuous, smooth glucose curves from gestational weeks 30-32. The curve fitting procedure was based on B-splines basis functions and is described in Appendix A.

The functional multilevel model

When functional data like glucose curves are observed on two or more occasions for each individual, we apply a multilevel model for functional data to extract information [26,27], as described below. See Appendix A for details.

Assume that the individual, true blood glucose curve $\gamma_{iv}(t)$ for woman $i = 1,\ldots,884$ at visit $v = 1,2$ in the continuous time span from 0 to 120 minutes, $t \in [0,120]$, can be decomposed into fixed and random effects curves (Figure 2A), and expressed as a multilevel model of functional data

$$\gamma_{iv}(t) = \mu(t) + \eta_v(t) + X_i(t) + U_{iv}(t). \qquad (1)$$

Here the fixed effects curves are the overall mean glucose curve $\mu(t)$ (Figure 2Ai), and the mean visit-specific deviation from the overall mean curve, $\eta_v(t)$ (Figure 2Aii). Together, these terms constitute the visit-specific mean curve, $\mu(t) + \eta_v(t)$ (Figure 2Cii). The random effects curves are $X_i(t)$, the subject-specific deviation from the visit-specific mean curve (Figure 2Aiii), and $U_{iv}(t)$, the subject- and visit-specific deviation from the subject-specific mean curve (Figure 2Aiv).

Extracting common temporal characteristics: functional principal component (FPC) curves

An important task in FDA is to quantify the common characteristics of a set of curves. The common characteristics of the curves $X_i(t)$, $i = 1,...,884$ and of the curves $U_{iv}(t)$, $i = 1,...,884$ and $v = 1,2$ in expression (1) (Figure 2Aiii and 2Aiv) are found by functional

principal component analysis (FPCA). FPCA extracts FPC curves that describe characteristics associated with the largest proportions of the variation in the curves. See Appendix A for details. The FPC curves estimated in the multilevel FPCA may be interpreted and labelled according to the physiological information they exhibit.

Estimating FPC scores

In order to quantify each individual glucose curve's characteristics, we estimate individual scores for each FPC curve. A woman's FPC scores reflect how her individual curve trajectories at the two visits correspond to the common characteristics expressed by the FPC curves. Then we can study how glucose curve characteristics vary from woman to woman, and from visit to visit.

A woman's scores for the FPC curves of $X_i(t)$ quantify her subject-specific deviation from the visit-specific mean curve, i.e. the important characteristics of her glucose curves *across* visits (Figure 2Biii and 2Ciii). Her scores for the FPC curves of $U_{iv}(t)$ quantify her subject- and visit-specific deviation from her subject-specific mean curve, i.e. the characteristics of the residual variation *within* a visit (Figure 2Biv and 2Civ).

By combining equation (1) with the FPC curves and corresponding estimated FPC scores, an individual glucose curve can be expressed as the sum of the visit-specific mean, $\mu(t)+\eta_v(t)$, and a linear combination of a small number of the FPC curves for $X_i(t)$ and $U_{iv}(t)$ (Equation (3) in the Appendix [26,27].

FPC curves are often illustrated by plots showing how an individual curve differs from the mean curve if the FPC scores are high or low, rather than plots of the FPC curves directly.

Correlation between FPC scores and traditional "area under the curve" (AUC)

The traditional, simple summary measure AUC was calculated directly from the de-trended glucose measurements by the trapezoid method. Pearson's correlation coefficients (r) were used to compare AUC with FPC scores.

Functional information in regression analyses

The impact of glucose curve characteristics on the neonatal outcomes birth weight, percentage fat and C-peptide in cord blood were estimated using linear regression with FPC scores from the multilevel FPCA as explanatory variables. The interpretation of the effect estimates is

based on the physiological interpretation of the FPC scores. Adjusted effect estimates were found by multiple linear regression analyses with most FPC scores (the first subject- and visit-specific score at gestational weeks 14-16 was left out due to colinearity issues), early pregnancy BMI, age and parity as explanatory variables. The multivariable analyses involved stepwise variable selection procedures based on Akaike's information criterion (AIC), analyses of several models considered to be of importance, and considerations of physiological importance of the findings. Model diagnostics were thoroughly checked during the analysis. The final multiple models presented in the results section include only the variables identified by these procedures.

In supplementary analyses, we did additional adjustment for weight gain between weeks 14-16 and 30-32, and repeated the analyses of birth weight for the reduced samples where percentage fat and C-peptide were available.

Software

All analyses were performed in R 3.0.0. The estimation of FPC scores was done by the R2WinBUGS package that runs WinBUGS from R [32]. The technical details are given in Appendix A, and the implementation is given in the program code, which is available as supporting information SI1.

Results

Data description

Characteristics of the study sample at inclusion at gestational weeks 14-16, at gestational weeks 30-32 and at birth are shown in Table 1. Except from a significantly lower proportion of smokers in the study sample (p=0.01), no significant differences were found between the women and the neonates in the study sample, and those who were excluded ($0.41 \leq p \leq 1.00$). There was a small increase in fasting glucose and a large increase in fasting insulin from inclusion to weeks 30-32. The 2-h glucose levels were elevated in third trimester, and the prevalence of GDM increased from 0.3% at inclusion to 6% in third trimester.

Fitted curves

The smoothed glucose curves at gestational weeks 14-16 (Figure 3A) and 30-32 (Figure 3B) showed large variations between the women at both visits. Glucose values were higher, and it took longer time for postprandial glucose levels to get back to fasting levels at gestational weeks 30-32 than at weeks 14-16.

Common temporal characteristics: FPC curves and FPC scores

In the multilevel FPCA, the first two subject-specific FPCs explained 98% of the variation *across* visits, and the first three subject- and visit-specific FPCs explained 92% of the residual variation *within* visits. Further analyses were restricted to these FPC curves and the corresponding FPC scores ($FPC1^{subj}$ and $FPC2^{subj}$, and $FPC1^{14\text{-}16}$, $FPC2^{14\text{-}16}$, $FPC3^{14\text{-}16}$, $FPC1^{30\text{-}32}$, $FPC2^{30\text{-}32}$ and $FPC3^{30\text{-}32}$, respectively). The FPC1 and FPC2 curves had very similar temporal appearances across and within visits.

Figure 3C-J show how individual curves differ from the overall and visit-specific mean curves if the FPC^{subj}, $FPC^{14\text{-}16}$ and $FPC^{30\text{-}32}$ scores are high or low. The dominating curve characteristic for the variation *across* visits, $FPC1^{subj}$, accounting for 91% of this variation, was given the interpretation "general glucose level". Women with high $FPC1^{subj}$ scores had glucose curves above the overall mean, and a somewhat later postprandial peak, whereas women with low $FPC1^{subj}$ scores had glucose curves below the overall mean (Figure 3C).

The second most important curve characteristic across visits ($FPC2^{subj}$) was "timing of postprandial peak". Women with low $FPC2^{subj}$ scores had a clear early peak and low glucose values at the end of the OGTTs (Figure 3D). Women with high $FPC2^{subj}$ scores had a later

postprandial peak and high glucose values at the end of the OGTTs. This was seen in plots of individual curves from women with the lowest and highest $FPC2^{subj}$ scores (plots not shown).

The dominating curve characteristic for the residual variation *within* visits (accounting for 72% of this variation), was "general glucose level within visits", i.e. the general glucose level not accounted for by the general glucose level *across* visits. A woman with a high subject- and visit-specific $FPC1^{14\text{-}16}$ ($FPC1^{30\text{-}32}$) score had a glucose curve above the subject-specific mean at weeks 14-16 (30-32) (Figure 3E and 3H). An example of this is the upper curve in Figure 2Biv and 2Civ. Similarly, a woman with a low $FPC1^{14\text{-}16}$ ($FPC1^{30\text{-}32}$) score had a glucose curve below the subject-specific mean at weeks 14-16 (30-32). An example of this is the lower curve in Figure 2Biv and 2Civ.

The second most important curve characteristic for the variation within visits was "timing of postprandial peak within visits". A woman with a low $FPC2^{14\text{-}16}$ ($FPC2^{30\text{-}32}$) score had a clear early peak, and low glucose values at the end of this OGTT (Figure 3F and 3I). A woman with a high $FPC2^{14\text{-}16}$ ($FPC2^{30\text{-}32}$) score at weeks 14-16 (30-32) had a later postprandial peak, with high glucose values at the end of this OGTT (plots of individual curves not shown).

The third most important curve characteristic for the variation within visits was "oscillating glucose within visits". The glucose curves of women with high $FPC3^{14\text{-}16}$ or $FPC3^{30\text{-}32}$ scores had two postprandial peaks during the corresponding OGTT, whereas women with low $FPC3^{14\text{-}16}$ or $FPC3^{30\text{-}32}$ scores had only one glucose peak during the OGTT (Figure 3G and 3J). The characteristics "peak" and "oscillations" accounted for a smaller part of the variation *across* visits (7% and less than 2%), than *within* visits (15% and 8%).

Figure 4 exemplifies the relation between individual glucose curves and corresponding FPC scores. The cyan, blue and black curves are glucose curves from three women in the study with curves above the mean (grey curve) at both visits. Consequently, the $FPC1^{subj}$ scores were high. The women with the green, red and purple curves, had low $FPC1^{subj}$ scores. The woman with cyan curves and generally high glucose levels on both OGTTs, had a curve that was high above the mean at gestational weeks 14-16, but less so at gestational weeks 30-32. Thus, her $FPC1^{14\text{-}16}$ score was high, and her $FPC1^{30\text{-}32}$ score low.

There were strong correlations between $FPC1^{subj}$ scores and AUC at weeks 14-16 and weeks 30-32 (r = 0.86 and 0.90, respectively), between. $FPC1^{14\text{-}16}$ and $AUC^{14\text{-}16}$ (r =0.73), and between $FPC1^{30\text{-}32}$ and $AUC^{30\text{-}32}$ (r= 0.87). See supporting information SI2 for a table of correlations. All FPC1 scores were positively correlated with BMI ($0.12 \leq r \leq 0.35$).

Regression analyses

Crude analyses showed significant positive effects on the three neonatal outcomes of the general glucose level across visits ($FPC1^{subj}$), and of the subject- and visit-specific general glucose level at weeks 30-32 ($FPC1^{30-32}$) (Table 2). There was also a significant effect of a subject-specific late peak ($FPC2^{subj}$) on birth weight.

In multivariable analyses of birth weight, only $FPC2^{sub}$and $FPC1^{30-32}$ scores remained significant (Table 2): Women with late postprandial peaks (low $FPC2^{subj}$ scores) would be expected to have babies with higher birth weight than women with early postprandial peaks (high $FPC2^{subj}$ scores), and women with the highest residual glucose levels in third trimester (high $FPC1^{30-32}$ scores) would be expected to have babies with higher birth weight than women with low residual glucose levels in third trimester (low $FPC1^{30-32}$ scores).

In multivariable analyses of neonatal percentage of fat, the effect of the $FPC1^{subj}$ scores remained significant (Table 2): Women with generally high glucose levels during their OGTTs (high $FPC1^{subj}$ scores) would be expected to have babies with a higher percentage of fat than women with generally low glucose levels during their OGTTs (low $FPC1^{subj}$ scores). According to AIC, $FPC2^{subj}$ scores also held information important for this outcome, although not statistically significant in the final model: Low $FPC2^{subj}$ scores, implying late glucose peaks, corresponded with high values of neonatal percentage fat.

In multivariable analyses of C-peptide in cord blood, the effect of $FPC1^{30-32}$ scores remained significant (Table 2): Women with high $FPC1^{30-32}$ scores gave birth to babies with higher mean C-peptide than women with low $FPC1^{30-32}$ scores. According to AIC, $FPC3^{14-16}$ scores also held important information, although not statistically significant: The neonates of women with oscillating glucose curves (high $FPC3^{14-16}$ scores) had somewhat lower C-peptide levels than those with one glucose peak during the OGTT (low $FPC3^{14-16}$ scores).

Supplementary analyses showed that alternative models chosen due to physiological theories, and to explore the effects of colinearity, gave the same results for all three outcomes. Additional adjustment for weight gain between the two visits did not change the final model for percentage fat or C-peptide, but replaced $FPC2^{subj}$ in the model for birth weight. Analyses of birth weight in reduced samples where percentage fat or C-peptide was available showed significant effects of $FPC1^{subj}$ in both subsamples, of $FPC2^{14-16}$ in the percentage fat subsample, and a non-significant contribution of $FPC2^{30-32}$ in the C-peptide subsample.

Discussion

The present study successfully used multilevel FDA to analyse changes in longitudinally observed glucose curves during pregnancy. The general glucose levels, in particular postprandial glucose, increased from early pregnancy to gestational weeks 30-32, and postprandial glucose peaked later in gestational weeks 30-32. The glucose characteristics extracted by FPCA had significant impact of glucose curve characteristics on birth weight, neonatal percentage of fat, and C-peptide in cord blood, demonstrating physiological relevance.

The physiological interpretation of FPC curves is essential for the usefulness of FPCA. The identification of the general glucose levels as the most dominant characteristics of individual glucose curves was supported by the strong associations between FPC1 scores and the AUCs. The elevated postprandial levels in third trimester, the small increase in fasting glucose, and the large increase in fasting insulin and prevalence of GDM from inclusion to weeks 30-32 (Table 1), are in accordance with an expected progressive insulin resistance among pregnant women [13]. This supports the finding of timing of postprandial peak as the second most important curve characteristic. The general glucose level accounted for a larger part of the variation across, than within visits, whereas the timing of the peak was more important for the variation within, than across visits. This is not surprising, as FPC2 represent more curvature than FPC1, and some of the curvature may be averaged out at the subject-specific level. The FPC3 curve was interpreted as "oscillating glucose within visits", although the scarcely sampled glucose measurements at each OGTT could only reveal two glucose peaks. The term "oscillations" was chosen due to physiological theories [33,34].

The regression results for birth weight in Table 2 were strengthened by the consistency of the results from alternative models. The lack of significant effect of $FPC2^{subj}$ after additional adjustment for weight gain is a questionable result, as weight gain may be on the causal pathway between the subject-specific glucose characteristics and birth weight, i.e. be an intermediate factor by which the mechanisms work [35]. This finding is therefore presented as a supplementary analysis in the text, but the main focus is on the results in Table 2. The different results for birth weight in the subsamples with percentage fat or cord blood C-peptide may be a consequence of colinearity issues in combination with the reduced sample size. Notably, in all these analyses, both a "general level" and a "timing of peak" characteristic were identified as important for birth weight.

The sample size in the present study was substantial, but the women were healthy and relatively homogenous. This may have caused less variation in the individual glucose curves

and made it more difficult to extract important discriminating curve characteristics. Also, the scarce sampling of glucose during the OGTT is likely to obscure the extraction and interpretation of the curve characteristics. More physiologically interesting temporal details and better discriminating abilities of the FPCs may be expected in a more heterogeneous population, and from OGTT curves over more than 2 hours or with a more frequent OGTT sampling. With more measurements per OGTT, it is also possible to apply alternative smoothing strategies [27]. In this study, the smoothing involved both a roughness penalty when fitting individual curves, and leaving out the FPCs which explained the smallest part of the variation, i.e. those with the waviest appearance. This might have given a too conservative estimate of the amount of curvature in the individual curves, which again could have caused bias in the FPC scores and thereby affected the regression results. It is also possible that the methods of covariance matrix estimation did not perfectly separate the across and within variances, influencing the colinearity, and thereby the variable selection.

Compared to studies presenting intravenous glucose tolerance tests [36], the glucose measurements per woman per OGTT in our study were few. They were, nevertheless, samples from an underlying, continuous and temporal process, and this made FDA a natural choice of analysis [24]. Alternative analyses include ordinary principal component analysis (PCA) of the five glucose measurements, and using these PCA scores as input in the regression analyses instead of FPCA scores. With only five measurements per curve, and measurements taken at the same time points for each woman, this would be expected to extract similar information as the FDA. However, FDA emphasizes the basic assumption about continuity of the underlying process, provides interpretations of curve features in this context, and opens for analysis of the derivatives of the curves [24]. FDA is also easier to apply in situations with more frequent sampling, sampling at unequal time points, and missing data.

The finding of the general glucose level as the most important glucose curve characteristic is in accordance with the numerous studies focusing on elevated glucose of various types, e.g. fasting, 1-h, 2-h or HbA1c values, in diabetes research [1,13,37]. Also, the increase in postprandial values during pregnancy, and corresponding delay in postprandial peak, found in the present study, is supported by several earlier studies [9,14,16,18,38]. FPC1 scores were positively associated with BMI, indicating that higher BMI leads to generally higher glucose levels. This is also in accordance with physiological knowledge of obesity and insulin resistance [13,39] and our recent findings [20]. Many studies have found a decline in fasting glucose during the first trimester of pregnancy [15], but an overview of longitudinal studies during pregnancy showed conflicting results concerning later pregnancy fasting

glucose [15]. This justifies our findings of a small increase in fasting glucose from weeks 14-16, to 30-32. Hence, current knowledge of metabolic changes during pregnancy supports the interpretations of the FPCs as plausible and potentially important physiological characteristics.

The HAPO study is a reference study for the impact of maternal blood glucose levels on pregnancy outcomes, and has found statistically significant higher odds ratios for high birth weight, cord-blood serum C-peptide level and percentage body fat (above their respective 90th percentiles), for high fasting, 1-hour and 2-hour glucose levels [1,40]. This supports our findings of important impact of FPC1 scores, interpreted as "general glucose level", on these outcomes. Other studies with a similar scope, but smaller sample sizes are also in accordance with these findings [2,4,6–8]. However, none of these studies addressed the impact of the dynamic regulation of the blood glucose, which is embedded in the FPC2 and FPC3 scores. Some studies have commented on the postprandial peak and birth outcomes [9,22,41], but to our knowledge, our previous study [20] is the only study that has formally investigated the impact of the timing of the postprandial peak.

We earlier found that for glucose curves from early pregnancy, the AUC was strongly correlated with FPC scores that provided information about the general glucose level during the OGTT, but not with scores providing information about timing of postprandial peak or oscillations, nor with the fasting or 2-h values [20]. This was also found in the present study (Supporting information SI2). Our recommendation is therefore to use AUC values rather than the fasting values or 2-h values, if FDA is not applied.

An important application of FDA techniques concerns research and clinical settings where continuous glucose monitoring devices are used [42]. Currently, many such studies restrict the analyses to simple summary measures like the mean glucose [17,43], resulting in loss of potentially important information, as demonstrated in [20]. With the increased use of continuous glucose monitoring, there is a strong need for methods that can extract important information from curve data.

The HAPO study extended the Pedersen hypothesis about how maternal hyperglycemia affects the foetus [13] to the normal-glycaemic range, thereby giving rise to a comprehensive debate about the GDM diagnosis [11,44–48]. In contrast to the WHO GDM criterion based on the fasting and 2-h value only [37], the new criteria suggested by the International Association of Diabetes and Pregnancy Study Group takes into consideration both the fasting, 1-h and 2-h OGTT values, with cut-off values based on risk estimates for adverse outcomes [48]. This implies that the new criteria indirectly address the dynamics in

the curves. We have shown that modern statistical analysis can extract curve information reflecting glucose dynamics that is important for both maternal [20] and neonatal outcomes. This can contribute to a better understanding of the different stages in the development of unhealthy glucose metabolism, and to a more precise prediction of women at risk for maternal or foetal complications. Then, interventions targeted to modify glucose curves could be initiated before a GDM diagnosis is given, or treatment for it is necessary. Such interventions have been studied in pregnant and non-pregnant study samples [49–51]. Future studies should investigate whether such interventions also may affect pregnancy outcomes, and have positive long-term effects on maternal health.

In conclusion, the physiologically interpretable glucose curve characteristics extracted by FDA in the present analysis, and their statistically significant effects of on birth weight, neonatal percentage fat, and cord blood C-peptide, show that shape information inherent in entire glucose curves is important for several outcomes, and may contribute to the understanding of the metabolic changes during pregnancy. FDA techniques can also be used to capture important curve information from more frequently sampled glucose curves, such as the increasingly used continuous glucose monitoring devices.

Appendix A: Multilevel functional data analysis (FDA)

The multilevel FDA was based on the works of Ramsay and Silverman [24], Di et al [26] and Crainiceanu and Goldsmith [27].

Fitting individually smoothed continuous curves

Let $\gamma_{iv}(t)$ be the underlying, true continuous and smooth glucose curve for woman $i = 1,\ldots,N$ at visit $v = 1,\ldots,V$. In our data, $N = 884$, $V = 2$ and $t \in [0,120]$. The estimation of individual curves $\hat{\gamma}_{iv}(t)$ from the observed discrete data points $y_{iv}(t_j)$, $j = 1,\ldots,J$, is based on the measurement model $y_{iv}(t_j) = \gamma_{iv}(t_j) + \varepsilon_{ivj}$, where $\gamma_{iv}(t_j)$ is γ_{iv} evaluated at time t_j and $\varepsilon_{ivj} \sim \mathrm{N}(0,\sigma^2)$ is an error term. In our data, $J = 5$. The individual curve estimates $\hat{\gamma}_{iv}(t)$ are found by subject-specific spline smoothing with B-splines basis functions and a roughness penalty [24].

The functional multilevel model

Assume that the individual continuous blood glucose curve $\gamma_{iv}(t)$ can be decomposed according to the model

$$\gamma_{iv}(t) = \mu(t) + \eta_v(t) + X_i(t) + U_{iv}(t). \qquad (2)$$

Here, $\mu(t)$ is the overall mean curve and $\eta_v(t)$ the mean visit-specific deviation from the overall mean curve, both assumed to be fixed effects curves. $X_i(t)$ and $U_{iv}(t)$ are assumed to be uncorrelated mean-zero stochastic processes, where $X_i(t)$ represent the subject-specific deviation from the visit-specific mean curve, $\mu(t) + \eta_v(t)$, and $U_{iv}(t)$ the subject- and visit-specific deviation from the subject-specific mean curve at visit *v*, $\mu(t) + \eta_v(t) + X_i(t)$.

Due to the large sample size, we assume that $\mu(t)$ can be estimated with negligible error by averaging the individual curve estimate $\hat{\gamma}_{iv}(t)$ over all subjects *i* and visits *v*,

$$\hat{\mu}(t) = \frac{1}{N \cdot V} \sum_{i,v} \hat{\gamma}_{iv}(t).$$

Similarly, we estimate the mean visit-specific deviation from the overall mean for visit v by

$$\hat{\eta}_v(t) = \left(\frac{1}{N} \sum_i \hat{\gamma}_{iv}(t) \right) - \hat{\mu}(t).$$

Extracting common temporal characteristics: functional principal component (FPC) curves

The common temporal characteristics for the stochastic processes $X_i(t)$ and $U_{iv}(t)$ can be found by functional principal component analysis (FPCA) of their temporal covariance surfaces K_X and K_U, for $t \in [0,120]$ and $s \in [0,120]$,

$$K_X(t,s) = \mathrm{cov}\left(X_i(t), X_i(s)\right)$$

and

$$K_U(t,s) = \mathrm{cov}\left(U_{iv}(t), U_{iv}(s)\right).$$

To obtain estimates of K_X and K_U, we first subtract the estimated visit-specific mean curve, $\mu(t) + \eta_v(t)$, from (2), giving

$$W_{iv}(t) = X_i(t) + U_{iv}(t),$$

i.e. the combined subject-specific, and subject- and visit-specific stochastic process. Let $K_W(t,s)$ be the covariance surface of $W_{iv}(t)$. Since $X_i(t)$ and $U_{iv}(t)$ are assumed to be uncorrelated this is simply

$$K_W(t,s) = \mathrm{cov}\left(W_{iv}(t), W_{iv}(s)\right) = K_X(t,s) + K_U(t,s).$$

Estimates of the covariance surfaces for $X_i(t)$ and $U_{iv}(t)$ can thus be based on $K_W(t,s)$. The only contribution to the between-visits covariance, $K_{W_{v \neq v'}}(t,s) = \mathrm{cov}\left(W_{iv}(t), W_{iv'}(s)\right)$, $v \neq v'$, is the subject-specific variation from $X_i(t)$, since $K_U(t,s) = \mathrm{cov}\left(U_{iv}(t), U_{iv'}(s)\right) = 0$ for $v \neq v'$ when assuming $\mathrm{cov}\left(X_i(t), U_{iv}(s)\right) = 0$. Hence, for $v \neq v'$,

$$K_{W_{v \neq v'}}(t,s) = \mathrm{cov}\left(W_{iv}(t), W_{iv'}(s)\right) = \mathrm{cov}\left(X_i(t), X_i(s)\right) = K_X(t,s).$$

That is, for $v \neq v'$, $K_X(t,s)$ can be estimated using a methods of moments estimator of $K_W(t,s)$. We denote this $\hat{K}_X(t,s)$. Consequently, for $v = v'$, $K_U(t,s)$ can be estimated by a corresponding methods of moments estimator

$$\hat{K}_U(t,s) = \hat{K}_{W_{v=v'}}(t,s) - \hat{K}_X(t,s).$$

Once $\hat{K}_X(t,s)$ and $\hat{K}_U(t,s)$ are available, the common temporal characteristics of $X_i(t)$ and $U_{iv}(t)$ can be estimated by principal component analysis of $\hat{K}_X(t,s)$ and $\hat{K}_U(t,s)$ as described in [26] and [27], extracting their corresponding functional principal component

(FPC) curves $\psi_a^X(t),\ a=1,\ldots,A$ and $\psi_b^U(t),\ b=1,\ldots,B.$ The FPC curves $\psi_a^X(t),$ $a=1,\ldots,A$ represent independent parts of the subject-specific temporal variability, and the FPC curves $\psi_b^U(t),\ b=1,\ldots,B$ represent independent parts of the subject- and visit-specific temporal variability.

Estimating FPC scores

The individual stochastic process $W_{iv}(t)$ can be expressed as linear combination of the FPC curves for $X_i(t)$ and $U_{iv}(t),$ by estimation of corresponding FPC scores. By retaining to the first A' principal component curves for $X_i(t)$, and the first B' principal component curves for $U_{iv}(t)$, the functional model for the joint stochastic process $W_{iv}(t)$ can be expressed as

$$W_{iv}(t)=\sum_{a=1}^{A'}\xi_{ia}\psi_a^X(t)+\sum_{b=1}^{B'}\zeta_{ivb}\psi_b^U(t)+\varepsilon_w(t); \qquad (3)$$

$$\xi_{ia}\sim N\left(0,\ \lambda_a^X\right);\ \zeta_{ivb}\sim N\left(0,\ \lambda_b^U\right);\ \varepsilon_w\sim N\left(0,\ \sigma_w^2\right),$$

where ξ_{ia} are the scores for $\psi_a^X(t),\ a=1,\ldots,A'$, ζ_{ivb} are the scores for $\psi_b^U(t),\ b=1,\ldots,B'$, λ_a^X is the eigenvalue of $\psi_a^X(t),$ λ_b^U is the eigenvalue of $\psi_b^U(t)$, and ε_w is an error term.

Implementation

The model in equation (3) can be implemented in WinBUGS to obtain estimates and corresponding estimation error, i.e. posterior distributions, of ξ_{ia} and ζ_{ivb} [27]. To completely specify the Bayesian model in WinBUGS, it is necessary to provide priors for all model parameters. Independent gamma priors with large dispersion were chosen as priors for all dispersion parameters in the model. The program code is available as supporting information SI1. The matrix notation needed for the implementation can be found in [26,27].

Individual curve estimates expressed by FPC curves and FPC scores

Combining equations (2) and (3) with estimated FPC scores $\hat{\xi}_{ia}$ and $\hat{\zeta}_{ivb}$, an individual glucose curve estimate $\tilde{\gamma}_{iv}(t)$ can be given as

$$\tilde{\gamma}_{iv}(t)=\hat{\mu}(t)+\hat{\eta}_v(t)+\sum_{a=1}^{A'}\hat{\xi}_{ia}\psi_a^X(t)+\sum_{b=1}^{B'}\hat{\zeta}_{ivb}\psi_b^U(t). \quad (4)$$

The glucose curve estimates $\hat{\gamma}_{iv}(t)$ from the penalised B-splines smoothing, and $\tilde{\gamma}_{iv}(t)$ from (4), will be similar, but not identical, as the linear combination in (4) is restricted to the first (A', B') FPC curves and scores. This restriction constitutes an additional smoothing of the observed glucose measurements.

Acknowledgements

We gratefully acknowledge the help of each woman who participated and all the professional personnel at Oslo University Hospital Rikshospitalet who made this study possible.

References

1. Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, Hadden DR, McCance DR, Hod M, McIntyre HD, Oats JJN, Persson B, Rogers MS, Sacks DA (2008) Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 358: 1991-2002.

2. Ferrara A, Weiss NS, Hedderson MM, Quesenberry CP, Selby JV, Ergas IJ, Peng T, Escobar GJ, Pettitt DJ, Sacks DA (2007) Pregnancy plasma glucose levels exceeding the American Diabetes Association thresholds, but below the National Diabetes Data Group thresholds for gestational diabetes mellitus, are related to the risk of neonatal macrosomia, hypoglycaemia and hyperbilirubinaemia. Diabetologia 50: 298-306.

3. Stotland NE, Caughey AB, Breed EM, Escobar GJ (2004) Risk factors and obstetric complications associated with macrosomia. Int J Gynaecol Obstet 87: 220-226.

4. McGowan CA, McAuliffe FM (2010) The influence of maternal glycaemia and dietary glycaemic index on pregnancy outcome in healthy mothers. Br J Nutr 104: 153-159.

5. Young BC, Ecker JL (2013) Fetal Macrosomia and Shoulder Dystocia in Women with Gestational Diabetes: Risks Amenable to Treatment? Curr Diab Rep 13: 12-18.

6. Scholl TO, Sowers M, Chen X, Lenders C (2001) Maternal glucose concentration influences fetal growth, gestation, and pregnancy complications. Am J Epidemiol 154: 514-520.

7. Ong KK, Diderholm B, Salzano G, Wingate D, Hughes IA, MacDougall J, Acerini CL, Dunger DB (2008) Pregnancy insulin, glucose, and BMI contribute to birth outcomes in nondiabetic mothers. Diabetes Care 31: 2193-2197.

8. Jeffery AN, Voss LD, Metcalf BS, Wilkin TJ (2004) The impact of pregnancy weight and glucose on the metabolic health of mother and child in the south west of the UK. Midwifery 20: 281-289.

9. Parretti E, Mecacci F, Papini M, Cioni R, Carignani L, Mignosa M, La TP, Mello G (2001) Third-trimester maternal glucose levels from diurnal profiles in nondiabetic pregnancies: correlation with sonographic parameters of fetal growth. Diabetes Care 24: 1319-1323.

10. Aparicio NJ, Joao MA, Cortelezzi M, Guz M, Sturgeon C, Galimberti DM, Fernandez CA (1998) Pregnant women with impaired tolerance to an oral glucose load in the afternoon: evidence suggesting that they behave metabolically as patients with gestational diabetes. Am J Obstet Gynecol 178: 1059-1066.

11. Wendland EM, Torloni MR, Falavigna M, Trujillo J, Dode MA, Campos MA, Duncan BB, Schmidt MI (2012) Gestational diabetes and pregnancy outcomes--a systematic review of the World Health Organization (WHO) and the International Association of Diabetes in Pregnancy Study Groups (IADPSG) diagnostic criteria. BMC Pregnancy Childbirth 12: 23.

12. Katon J, Williams MA, Reiber G, Miller E (2011) Antepartum A1C, maternal diabetes outcomes, and selected offspring outcomes: an epidemiological review. Paediatr Perinat Epidemiol 25: 265-276.

13. Hod M, Jovanovic L, Di Renzo GC, de Leiva A, Langer O (2008) Textbook of diabetes and pregnancy. Andover, Hampshire: Taylor & Francis. 628 p.

14. Hadden DR, McLaughlin C (2009) Normal and abnormal maternal metabolism during pregnancy. Semin Fetal Neonatal Med 14: 66-71.

15. Mills JL, Jovanovic L, Knopp R, Aarons J, Conley M, Park E, Lee YJ, Holmes L, Simpson JL, Metzger B (1998) Physiological reduction in fasting plasma glucose concentration in the first trimester of normal pregnancy: The diabetes in early pregnancy study. Metabolism 47: 1140-1144.

16. Lind T, Billewic WZ, Brown G (1973) Serial Study of Changes Occurring in Oral Glucose-Tolerance Test During Pregnancy. J Obstet Gynaecol Br Commonw 80: 1033-1039.
17. Siegmund T, Rad NT, Ritterath C, Siebert G, Henrich W, Buhling KJ (2008) Longitudinal changes in the continuous glucose profile measured by the CGMS in healthy pregnant women and determination of cut-off values. Eur J Obstet Gynecol Reprod Biol 139: 46-52.
18. Catalano PM, Tyzbir ED, Wolfe RR, Calles J, Roman NM, Amini SB, Sims EAH (1993) Carbohydrate-Metabolism During Pregnancy in Control Subjects and Women with Gestational Diabetes. Am J Physiol 264: E60-E67.
19. Hernandez TL, Friedman JE, Van Pelt RE, Barbour LA (2011) Patterns of glycemia in normal pregnancy: should the current therapeutic targets be challenged?. Diabetes Care 34: 1660-1668.
20. Frøslie KF, Røislien J, Qvigstad E, Godang K, Bollerslev J, Voldner N, Henriksen T, Veierød MB (2013) Shape information from glucose curves: Functional data analysis compared with traditional summary measures. BMC Med Res Methodol 13: 6.
21. Voldner N, Qvigstad E, Frøslie KF, Godang K, Henriksen T, Bollerslev J (2010) Increased risk of macrosomia among overweight women with high gestational rise in fasting glucose. J Matern Fetal Neonatal Med 23: 74-81.
22. Jovanovicpeterson L, Peterson CM, Reed GF, Metzger BE, Mills JL, Knopp RH, Aarons JH (1991) Maternal Postprandial Glucose-Levels and Infant Birth-Weight - the Diabetes in Early-Pregnancy Study. Am J Obstet Gynecol 164: 103-111.
23. Ramsay JO: Functional data analysis. Available: http://www.functionaldata.org. Accessed 15 October 2013.
24. Ramsay JO, Silverman BW (2005) Functional Data Analysis. New York: Springer. 426 p.
25. Voldner N, Frøslie KF, Bø K, Haakstad L, Hoff C, Godang K, Bollerslev J, Henriksen T (2008) Modifiable determinants of fetal macrosomia: role of lifestyle-related factors. Acta Obstet Gynecol Scand 87: 423-429.
26. Di CZ, Crainiceanu CM, Caffo BS, Punjabi NM (2009) Multilevel Functional Principal Component Analysis. Ann Appl Stat 3: 458-488.
27. Crainiceanu C, Goldsmith AJ (2010) Bayesian Functional Data Analysis Using WinBUGS. J Stat Softw 32: 1-33.
28. Voldner N (2010) Modifiable determinants of newborn macrosomia and birth complications. Dissertation for the Degree of PhD. Oslo: University of Oslo, Faculty of Medicine. 63 p.
29. Frøslie KF, Godang K, Bollerslev J, Henriksen T, Røislien J, Veierød MB, Qvigstad E (2011) Correction of an unexpected increasing trend in glucose measurements during 7years recruitment to a cohort study. Clin Biochem 44: 1483-1486.
30. Godang K, Qvigstad E, Voldner N, Isaksen GA, Frøslie KF, Nøtthellen J, Henriksen T, Bollerslev J (2010) Assessing body composition in healthy newborn infants: reliability of dual-energy x-ray absorptiometry. J Clin Densitom 13: 151-160.
31. Godang K, Frøslie KF, Henriksen T, Isaksen GA, Voldner N, Lekva T, Ueland T, Bollerslev J (2013) Umbilical cord levels of sclerostin, placental weight, and birth weight are predictors of total bone mineral content in neonates. Eur J Endocrinol 168: 371-378.
32. The R Foundation for Statistical Computing: R verison 3.0.2 (2013-09-25). Available: http://www.r-project.org. Accessed 01 October 2013.

33. Trujillo-Arriaga HM, Roman-Ramos R (2008) Fitting and evaluating the glucose curve during a quasi continuous sampled oral glucose tolerance test. Comput Biol Med 38: 185-195.
34. Li J, Kuang Y, Mason CC (2006) Modeling the glucose-insulin regulatory system and ultradian insulin secretory oscillations with two explicit time delays. J Theor Biol 242: 722-735.
35. Friis CM, Frøslie KF, Røislien J, Voldner N, Godang K, Ueland T, Bollerslev J, Veierød MB, Henriksen T (2010) The interleukins IL-6 and IL-1Ra: a mediating role in the associations between BMI and birth weight? J Dev Orig Health Dis 1: 310-318.
36. Sivan E, Chen X, Homko CJ, Reece EA, Boden G (1997) Longitudinal study of carbohydrate metabolism in healthy obese pregnant women. Diabetes Care 20: 1470-1475.
37. Alberti KGMM, Zimmet PZ (1998) Definition, diagnosis and classification of diabetes mellitus and its complications part 1: Diagnosis and classification of diabetes mellitus - Provisional report of a WHO consultation. Diabet Med 15: 539-553.
38. Cousins L, Rigg L, Hollingsworth D, Brink G, Aurand J, Yen SSC (1980) 24-Hour Excursion and Diurnal Rhythm of Glucose, Insulin, and C-Peptide in Normal-Pregnancy. Am J Obstet Gynecol 136: 483-488.
39. World Health Organization (2006) Definition and diagnosis of diabetes mellitus and intermediate hypergycemia. Report of a WHO/IDF consultation. Geneva, Switzerland: World Health Organization. 46 p.
40. HAPO Study Cooperative Research Group (2009) Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study: associations with neonatal anthropometrics. Diabetes 58: 453-459.
41. Beardsall K, Diderholm BMS, Dunger DB (2008) Insulin and carbohydrate metabolism. Best Pract Res Clin Endocrinol Metab 22: 41-55.
42. Chitayat L, Zisser H, Jovanovic L (2009) Continuous Glucose Monitoring During Pregnancy. Diabetes Technol Ther 11: S105-S111.
43. Buhling KJ, Winkel T, Wolf C, Kurzidim B, Mahmoudi M, Wohlfarth K, Wascher C, Schink T, Dudenhausen JW (2005) Optimal timing for postprandial glucose measurement in pregnant women with diabetes and a non-diabetic pregnant population evaluated by the Continuous Glucose Monitoring System (CGMS (R)). J Perinat Med 33: 125-131.
44. Coustan DR (2012) Point: the American Diabetes Association and the International Association of Diabetes and Pregnancy study groups recommendations for diagnosing gestational diabetes should be used worldwide. Clin Chem 58: 1094-1097.
45. Langer O, Umans JG, Miodovnik M (2013) The proposed GDM diagnostic criteria: a difference, to be a difference, must make a difference. J Matern Fetal Neonatal Med 26: 111-115.
46. Long H, Cundy T (2013) Establishing consensus in the diagnosis of gestational diabetes following HAPO: where do we stand? Curr Diab Rep 13: 43-50.
47. Coustan DR, Lowe LP, Metzger BE, Dyer AR (2010) The Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study: paving the way for new diagnostic criteria for gestational diabetes mellitus. Am J Obstet Gynecol 202: 654-656.
48. Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, Damm P, Dyer AR, Leiva A, Hod M, Kitzmiler JL, Lowe LP, McIntyre HD, Oats JJ, Omori Y, Schmidt MI (2010) International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 33: 676-682.

49. Gumbiner B, Van CE, Beltz WF, Ditzler TM, Griver K, Polonsky KS, Henry RR (1996) Abnormalities of insulin pulsatility and glucose oscillations during meals in obese noninsulin-dependent diabetic patients: effects of weight reduction. J Clin Endocrinol Metab 81: 2061-2068.
50. Clapp JF, III (1998) Effect of dietary carbohydrate on the glucose and insulin response to mixed caloric intake and exercise in both nonpregnant and pregnant women. Diabetes Care 21 Suppl 2: B107-B112.
51. Dunstan DW, Kingwell BA, Larsen R, Healy GN, Cerin E, Hamilton MT, Shaw JE, Bertovic DA, Zimmet PZ, Salmon J, Owen N (2012) Breaking up prolonged sitting reduces postprandial glucose and insulin responses. Diabetes Care 35: 976-983.

Figure legends

Figure 1: Flow chart

Figure 2: The functional multilevel model.

In all plots in the figure, the horizontal axis is time during the 2 hour oral glucose tolerance test, and the vertical axis is blood glucose, with range from -2 to 12.5 mmol/l. The horizontal, grey line is 0 mmol/l. $\gamma_{iv}(t)$ is the glucose curve from 0 to 120 min for woman $i = 1,\ldots,884$ at visit $v = 1,2$; $\mu(t)$ is the overall mean glucose curve; $\eta_v(t)$ is the visit-specific deviation from the overall mean curve; $X_i(t)$ is the subject-specific deviation from the visit-specific mean curve; and $U_{iv}(t)$ is the subject- and visit-specific deviation from the subject-specific mean curve.

Figure 3: Smoothed glucose curves and results from the multilevel FPCA.

Plots A and B show individually smoothed curves from gestational weeks 14-16 and 30-32 (black lines) and the visit-specific mean curves (grey lines).

Plots C and D show the overall mean of the fitted curves (grey lines) and how the shape of an individual curve differs from the overall mean if a multiplum of the subject-specific FPC curves is added to (+) or subtracted from (-) the mean curve.

Plots E-G and H-J show the visit-specific means at gestational weeks 14-16 and 30-32, respectively (grey lines), and how the shape of an individual curve differs from the visit-specific mean if a multiplum of the subject- and visit-specific FPC curves is added to (+) or subtracted from (-) this mean.

The multiplums correspond to 2 SDs of the corresponding FPC scores.

Figure 4: Examples of individual curves and corresponding scores.

The upper, left plot shows the individual glucose curves from 6 women at gestational weeks 14-16, and the upper, right plot shows the glucose curves from the same 6 women at gestational weeks 30-32. The lower plot shows the FPC scores for the same 6 women. The grey curves in the upper plots are the mean glucose curves at gestational weeks 14-16 (left) and 30-32 (right). Correspondingly, the grey line in the lower plot is zero.

Table 1: Sample characteristics.

		Study sample, n=884[*]		Excluded, n=90[*]
			Range	
	Maternal age (years)	31 (4)	19 – 42	31 (4)
	Para 0	461 (52%)		43 (48%)
	Daily smoker[†]	15 (2%)		6 (7%)
	Height (cm)	169 (6)	150 – 184	169 (6)
Inclusion	Gestational weeks	15.8 (1.3)	12.1 – 22.0	15.8 (1.4)
	Weight (kg)	69.8 (12.1)	44.6 – 123.1	70.3 (11.4)
	Body mass index (kg/m^2)	24.5 (4.0)	17.2 – 44.0	24.4 (3.5)
	Fasting blood glucose (mmol/l)	4.0 (0.4)	2.6 – 5.3	
	120 min blood glucose (mmol/l)	4.1 (1.1)	1.2 – 7.8	
	Fasting insulin (pmol/l) median (Q_1,Q_3)	27 (18, 39)	8 – 305	
	GDM[‡]	3 (0.3%)		
Third trimester	Gestational week	31.2 (1.0)	26.0 – 35.4	31.2 (0.8)
	Weight gain from inclusion	7.7 (2.6)	-2 – 22	7.8 (2.1)
	Fasting blood glucose (mmol/l)	4.1 (0.5)	3.0 – 6.2	
	120 min blood glucose (mmol/l)	5.5 (1.3)	1.9 – 10.3	
	Fasting insulin (pmol/l) median (Q_1,Q_3)	41 (26, 61)	8 – 989	
	GDM[‡]	50 (6%)		
Birth	Gestational week	40.2 (1.2)	37.0 – 43.1	40.1 (1.2)
	Birth weight child (g)	3654 (481)	2315 – 5420	3697 (527)
	Total % fat[§] (n=187)	13.6 (2.4)	8 – 20	13.2 (2.1)
	C-peptide in cord blood (ng/ml) (n=137)	1.1 (0.7)	0.1 – 5.0	1.1 (0.9)

Characteristics of the study sample and those excluded due to incomplete OGTT data. Results are presented as means (SDs) for continuous variables and frequencies (%) for categorical variables, unless otherwise stated.

[*] n may vary due to missing values.

[†] More than 1 cigarette/day.

[‡] GDM; gestational diabetes: 120 min glucose at or above 7.8 mmol/l

[§] Percentage fat estimated by DXA scan

Table 2: Regression analyses.

Outcome:		Birth weight, n=868			Percentage fat of newborn, as measured by DXA, n=185			C-peptide in cord blood, n=134		
Crude estimates		B	95% CI	p	B	95% CI	p	B	95% CI	p
Subject-specific Wks 14-16	$FPC1^{subj}$, "Level"	51	(30,73)	<0.001	0.34	(0.11,0.57)	<0.001	0.08	(0.00,0.16)	0.05
	$FPC2^{subj}$, "Peak"	-112	(-205,-18)	0.02	-1.10	(-2.26,0.06)	0.06	-0.11	(-0.53,0.31)	0.61
	$FPC1^{14-16}$, "Level"	-21	(-50,8)	0.16	-0.02	(-0.35,0.31)	0.91	-0.09	(-0.21,0.02)	0.11
	$FPC2^{14-16}$, "Peak"	-23	(-94,47)	0.52	-0.37	(-1.12,0.38)	0.33	-0.12	(-0.39,0.15)	0.38
	$FPC3^{14-16}$, "Oscill"	28	(-68,124)	0.56	0.43	(-0.64,1.49)	0.43	-0.30	(-0.69,0.08)	0.12
Wks 30-32	$FPC1^{30-32}$, "Level"	51	(26,76)	<0.001	0.26	(0.00,0.52)	0.05	0.14	(0.05,0.23)	<0.001
	$FPC2^{30-32}$, "Peak"	-29	(-87,28)	0.32	-0.16	(-0.89,0.56)	0.66	0.07	(-0.20,0.34)	0.60
	$FPC3^{30-32}$, "Oscill"	-20	(-101,60)	0.62	-0.66	(-1.59,0.27)	0.16	-0.16	(-0.47,0.16)	0.32
Adjusted estimates from multivariable analyses*										
Subject-specific Wks 14-16	$FPC1^{subj}$, "Level"				0.25	(0.01,0.49)	0.04			
	$FPC2^{subj}$, "Peak"	-92	(-181,-3)	0.04	-0.89	(-2.03,0.26)	0.13			
	$FPC1^{14-16}$, "Level"									
	$FPC2^{14-16}$, "Peak"									
	$FPC3^{14-16}$, "Oscill"							-0.34	(-0.72,0.03)	0.07
Wks 30-32	$FPC1^{30-32}$, "Level"	38	(14,62)	0.002				0.13	(0.04,0.23)	0.004
	$FPC2^{30-32}$, "Peak"									
	$FPC3^{30-32}$, "Oscill"									
BMI	Wks 14-16	24	(16,32)	<0.001	0.09	(-0.01,0.18)	0.06	0.02	(-0.01,0.05)	0.13
Parity		189	(128,251)	<0.001						

Results from univariable and multivariable regression analyses, with birth weight, neonatal percentage fat, or C-peptide in cord blood as response variables. * Multivariable analyses included all glucose variables except $FPC1^{14-16}$, due to colinearity diagnostics. Variable selection was done by Akaike's information criterion.

Figure 1

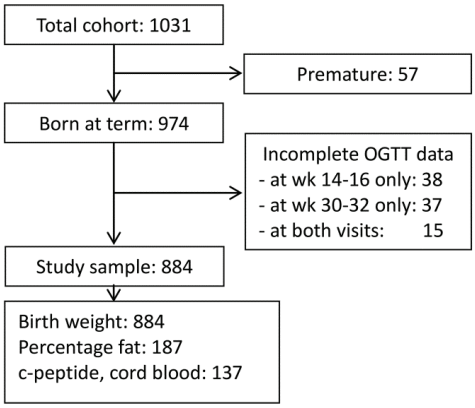
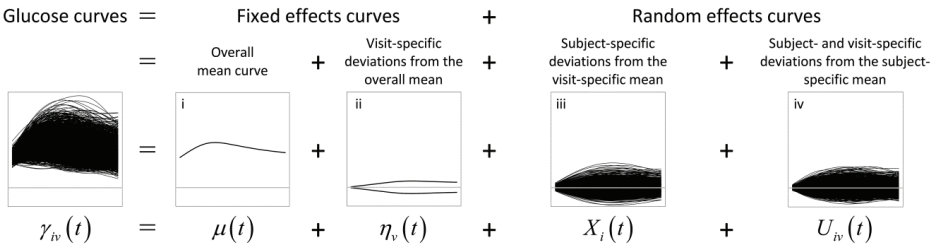
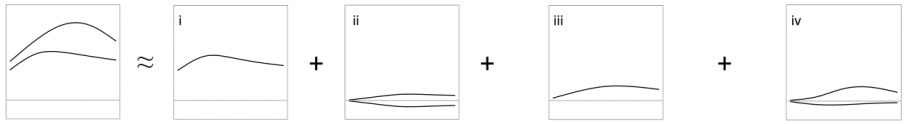

Total cohort: 1031
Premature: 57
Born at term: 974
Incomplete OGTT data
- at wk 14-16 only: 38
- at wk 30-32 only: 37
- at both visits: 15
Study sample: 884
Birth weight: 884
Percentage fat: 187
c-peptide, cord blood: 137

Figure 2

A The functional multilevel model



B An example: The estimated decomposition of glucose curves for woman no. 828



C Example (continued): Building up the glucose curve estimates for woman no. 828

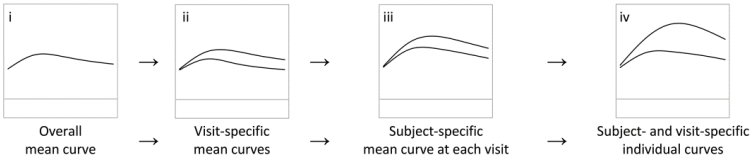


Figure 3

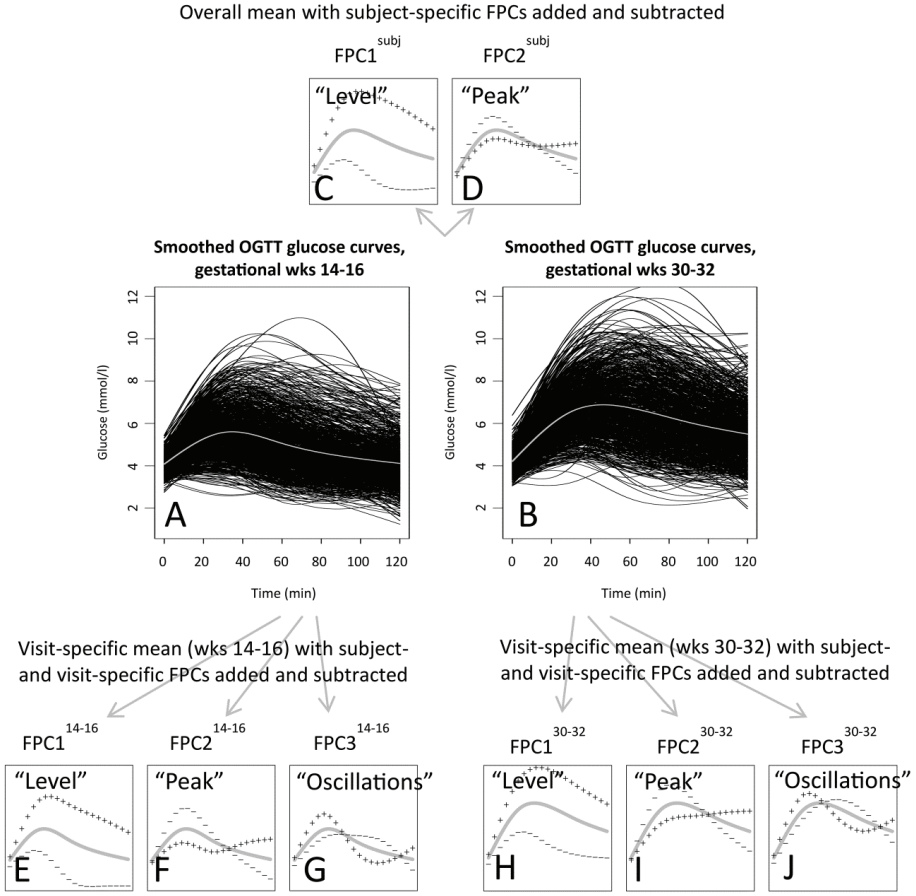
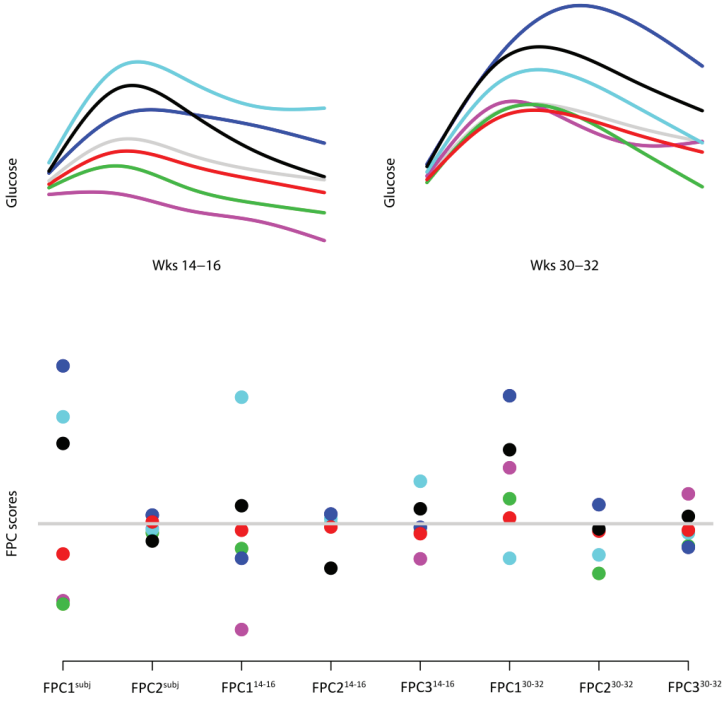

Overall mean with subject-specific FPCs added and subtracted
FPC1 subj
FPC2 subj
"Level"
"Peak"
C
D
Smoothed OGTT glucose curves, gestational wks 14-16
Smoothed OGTT glucose curves, gestational wks 30-32
Glucose (mmol/l)
Time (min)
A
B
Visit-specific mean (wks 14-16) with subject- and visit-specific FPCs added and subtracted
Visit-specific mean (wks 30-32) with subject- and visit-specific FPCs added and subtracted
FPC1 14-16
FPC2 14-16
FPC3 14-16
FPC1 30-32
FPC2 30-32
FPC3 30-32
"Level"
"Peak"
"Oscillations"
E
F
G
H
I
J

Figure 4


Glucose
Wks 14–16
Glucose
Wks 30–32
FPC scores
FPC1subj
FPC2subj
FPC114-16
FPC214-16
FPC314-16
FPC130-32
FPC230-32
FPC330-32

Supplementary material, Article IV

Correlations between glucose measurements and functional principal component scores in
Shape information in repeated glucose curves during pregnancy provided significant physiological information for neonatal outcomes, by Frøslie et al.

		OGTT, wks 14-16					OGTT, wks 30-32														
		Fasting	30 min	60 min	90 min	120 min	Fasting	30 min	60 min	90 min	120 min	AUC^{14-16}	AUC^{30-32}	$FPC1^{subj}$ scores	$FPC2^{subj}$ scores	$FPC1^{14-16}$ scores	$FPC2^{14-16}$ scores	$FPC3^{14-16}$ scores	$FPC1^{30-32}$ scores	$FPC2^{30-32}$ scores	$FPC3^{30-32}$ scores
OGTT, wks 14-16	Fasting	1.00																			
	30 min	0.55	1.00																		
	60 min	0.45	0.87	1.00																	
	90 min	0.43	0.72	0.90	1.00																
	120 min	0.42	0.58	0.71	0.83	1.00															
OGTT, wks 30-32	Fasting	0.47	0.34	0.32	0.31	0.28	1.00														
	30 min	0.35	0.49	0.45	0.41	0.27	0.57	1.00													
	60 min	0.33	0.52	0.54	0.51	0.37	0.49	0.85	1.00												
	90 min	0.31	0.50	0.56	0.55	0.44	0.47	0.70	0.91	1.00											
	120 min	0.27	0.38	0.45	0.50	0.48	0.44	0.50	0.65	0.81	1.00										
AUC^{14-16}		0.55	0.90	0.97	0.94	0.81	0.36	0.47	0.55	0.57	0.49	1.00									
AUC^{30-32}		0.36	0.53	0.56	0.54	0.42	0.58	0.87	0.97	0.95	0.78	0.58	1.00								
$FPC1^{subj}$ scores		0.54	0.77	0.83	0.82	0.69	0.58	0.75	0.86	0.88	0.74	0.86	0.90	1.00							
$FPC2^{subj}$ scores		-0.21	-0.29	-0.08	0.13	0.39	-0.19	-0.41	-0.17	0.07	0.45	-0.03	-0.09	-0.03	1.00						
$FPC1^{14-16}$ scores		0.24	0.61	0.72	0.70	0.66	-0.12	-0.15	-0.12	-0.09	-0.05	0.73	-0.12	0.29	0.09	1.00					
$FPC2^{14-16}$ scores		-0.08	-0.50	-0.32	0.02	0.30	0.01	-0.08	-0.13	-0.15	-0.15	-0.21	-0.14	-0.15	0.39	-0.09	1.00				
$FPC3^{14-16}$ scores		0.43	0.40	0.02	-0.07	0.31	0.18	0.12	0.05	0.03	0.05	0.16	0.07	0.13	-0.09	0.09	-0.08	1.00			
$FPC1^{30-32}$ scores		0.05	0.13	0.13	0.12	0.05	0.38	0.73	0.86	0.84	0.70	0.12	0.87	0.58	0.00	-0.54	-0.10	-0.04	1.00		
$FPC2^{30-32}$ scores		0.04	0.07	0.12	0.15	0.16	0.02	-0.35	-0.10	0.24	0.59	0.13	0.04	0.13	0.67	0.08	-0.30	-0.03	0.05	1.00	
$FPC3^{30-32}$ scores		0.10	-0.03	-0.10	-0.06	0.00	0.28	0.16	-0.24	-0.27	0.19	-0.06	-0.09	-0.07	0.07	-0.02	0.04	0.13	-0.11	0.06	1.00

```
###     Program code for the multilevel functional data analysis in
###     «Shape information in repeated glucose curves during pregnancy provided significant physiological information for neonatal outcomes.»,
###     by Frøslie et al.
###     Parts of the program code are clsosely based on the code in «Bayesian Functional Data Analysis Using WinBUGS»,
###     by Crainiceanu and Goldsmith, Journal of Statistical Software, 2010.


       # Read spss data from the complete cohort of 1031 women

library(foreign)
spssdata                      <- read.spss("M:/Art5longitudinalFDA/STORK29.06.2011.sav",use.value.labels = TRUE, to.data.frame = TRUE)
spssdata                      <- spssdata[order(spssdata$id),]
spssdata$auc1                 <- 0.5*(spssdata$g01/2+spssdata$g11+spssdata$g21+spssdata$g31+spssdata$g41/2)
spssdata$auc3                 <- 0.5*(spssdata$g03/2+spssdata$g13+spssdata$g23+spssdata$g33+spssdata$g43/2)

       # Remove women with premature births or non-complete OGTT data to get the sudy sample of 884 women and their neonates

spssdata.term.complete.ogtt   <- spssdata[spssdata$gest>=37&is.na(spssdata$gest)==FALSE&
                                          is.na(spssdata$g01*spssdata$g11*spssdata$g21*spssdata$g31*spssdata$g41*
                                                spssdata$g03*spssdata$g13*spssdata$g23*spssdata$g33*spssdata$g43)==FALSE,]

       # Make the 884x5-dimensional matrix of glucose data (raw data) from visit 1 and the one from visit 3

g1                            <- as.matrix(cbind(   spssdata.term.complete.ogtt$g01,spssdata.term.complete.ogtt$g11,

       spssdata.term.complete.ogtt$g21,spssdata.term.complete.ogtt$g31,spssdata.term.complete.ogtt$g41))
g3                            <- as.matrix(cbind(   spssdata.term.complete.ogtt$g03,spssdata.term.complete.ogtt$g13,

       spssdata.term.complete.ogtt$g23,spssdata.term.complete.ogtt$g33,spssdata.term.complete.ogtt$g43))

N_subj                        <- dim(g1)[1]         # 884
N_obs                         <- dim(g1)[2]         # 5
g                             <- rbind(g1,g3)       # dim: 1768,5
ng                            <- length(g[,1])

       # Plots. Remark: All plots in this code are not necessary, but may be useful for different purposes. The code is therefore kept in this file.
X11()
breaksuse                     <- c(0,30,60,90,120)
matplot(breaksuse,t(g[1:ng,]), type="l", lty=1, col="dark grey",lw=1,xlab="Time in minutes", ylab="Glucose in mmol/l")
X11()
par(mar=c(1,1,1,1))
matplot(breaksuse,t(g[1:ng,]), type="b", pch=20,lty=1,col="grey",lw=1,xaxt='n', yaxt='n',ann=FALSE,ylim=c(-0.5,12.5))
matplot(breaksuse,t(g[1:ng,]), col="black",pch=20,cex=2,add=TRUE)
abline(h=0,col="grey",lw=5)

       # Make B-spline basis. May have to install the fda package first: install.packages("fda")

library(fda)
mybasis                       <- create.bspline.basis(rangeval=c(0,120),norder=4, breaks=c(0,30,60,90,120))
```

```
       # Optimise lambda

#loglam                        <- seq(-30,20,0.05)
loglam                         <- seq(0,20,0.05)            # Example of alternativ range for loglam in case of local minimum (see below)
nlam                                <- length(loglam)
dfsave                         <- rep(NA,nlam)
gcvsave                        <- rep(NA,nlam)

for (ilam in 1:nlam)
{
Lambda                         <- 10^loglam[ilam]
fdParobj                       <- fdPar(mybasis,2,lambda)
smoothlist                     <- smooth.basis(c(0,30,60,90,120),t(g),fdParobj)
dfsave[ilam]                   <- smoothlist$df
gcvsave[ilam]                  <- sum(smoothlist$gcv)
}

       # Optimal value of lambda:

lambdaopt                      <- 10^loglam[gcvsave==min(gcvsave)]  #       lambdaopt <- 707.9458

       # Plots showing how gcv vary with loglambda.
       # (To ensute that the optimal value is not a local minimum. If so, choose an alternativ range for lambda (see above) )

X11()
par(mfrow=c(2,3))
plot(loglam,gcvsave,type="l")
plot(loglam,gcvsave,type="l",ylim=c(2200,4000),xlim=c(0,5))
plot(loglam,gcvsave,type="l",ylim=c(2300,2400),xlim=c(2.5,3.2))
abline(h=min(gcvsave),lty=2)
plot(loglam,gcvsave,type="l",ylim=c(2340,2360),xlim=c(2.7,3.1))
abline(h=min(gcvsave),lty=2)

       # Optimal smoothing of individual glucose curves, according to gcv criterion

fdParobj.opt                   <- fdPar(mybasis,2,lambdaopt)

       # Smoothed, individual curves (the basic units in further FDA), and corresponding evaluated function values

g1.smooth                      <- smooth.basis(c(0,30,60,90,120),t(g1),fdParobj.opt)      # These are the correct curves and glucose values
g3.smooth                      <- smooth.basis(c(0,30,60,90,120),t(g3),fdParobj.opt)      # for further analysis
eval.g1                        <- t(eval.fd(c(0,30,60,90,120),g1.smooth$fd))
eval.g3                        <- t(eval.fd(c(0,30,60,90,120),g3.smooth$fd))
eval.g                         <- rbind(eval.g1,eval.g3)
eval.error                     <- g-eval.g

       # Means of smoothed function values

overallmean.eval               <- colMeans(eval.g    ,na.rm=TRUE)
visitspecificmean1.eval        <- colMeans(eval.g1   ,na.rm=TRUE)
visitspecificmean3.eval        <- colMeans(eval.g3   ,na.rm=TRUE)
```

```
        # Minimal smoothing of mean curves (means of smoothed function values), to optain continuous mean curves

fdParobj.m                          <- fdPar(mybasis,2,1)

overallmean.eval.smooth             <- smooth.basis(c(0,30,60,90,120),overallmean.eval,fdParobj.m )
visitspecificmean1.eval.smooth      <- smooth.basis(c(0,30,60,90,120),visitspecificmean1.eval,fdParobj.m )
visitspecificmean3.eval.smooth      <- smooth.basis(c(0,30,60,90,120),visitspecificmean3.eval,fdParobj.m )


        # Plots
X11()
boxplot(eval.error,names=c("0","30","60","90","120"),ylim=c(-5.5,5.5),xaxt='n', yaxt='n',ann=FALSE)
X11()
par(mfrow=c(1,2))
plot(g1.smooth,lty=1,col="black",ylim=c(1,12),xlab="Time (min)",ylab="Glucose (mmol/l)",main="Smoothed OGTT glucose curves, \n gestational wks 14-16")
plot(visitspecificmean1.eval.smooth,lty=1,col="grey",lw=2,add=TRUE)
plot(g3.smooth,lty=1,col="black",ylim=c(1,12),xlab="Time (min)",ylab="Glucose (mmol/l)",main="Smoothed OGTT glucose curves, \n gestational wks 30-32")
plot(visitspecificmean3.eval.smooth,lty=1,col="grey",lw=2,add=TRUE)

        # Next step: Center the estimated visit 1 and 3 functional values on the visit-specific mean

g1.demeaned                         <- eval.g1 - matrix(rep(colMeans(eval.g1,na.rm=TRUE),nrow(eval.g1)),nrow=nrow(eval.g1),byrow=TRUE)
g3.demeaned                         <- eval.g3 - matrix(rep(colMeans(eval.g3,na.rm=TRUE),nrow(eval.g3)),nrow=nrow(eval.g3),byrow=TRUE)

        # Combine the de-meaned values in one matrix

g1and3.demeaned                     <- cbind(g1.demeaned,g3.demeaned)

        # Calculate the covariance matrix of the data centered at the visit-specific means.

big_covariance_v1and3               <- cov(g1and3.demeaned,use="pairwise.complete.obs")
round(big_covariance_v1and3,2)
big_correlation_v1and3              <- cor(g1and3.demeaned,use="pairwise.complete.obs")
round(big_correlation_v1and3,2)

        # Multilevel analysis       Gt: Gtotal     Gb: Gbetween   Gw: Gwithin
N                           <- N_obs
Gt                          <- (big_covariance_v1and3[1:N ,1: N]+big_covariance_v1and3[(N+1): (2*N),(N+1): (2*N)])/2
Gb                          <- (big_covariance_v1and3[1:N ,(N+1): (2*N)]+big_covariance_v1and3[(N+1): (2*N),1: N])/2 # This is the estimated Kx
Gw                          <- Gt-Gb                                                                                # This is the estimated Ku

        # Remark: No covariances need here to be smoothed (as compared to the works of Crainiceanu and Di),
        # as we have smoothed the curves as our data preparation step.


        # Plots of the estimated Ku and Kx surfaces
X11()
par(mfrow=c(1,2))
contour(seq(0,120,30),seq(0,120,30),matrix(Gw,nrow=5,byrow=TRUE),main="Gwds (Original Gt matrix with Gb subtr from all elements.)")
contour(seq(0,120,30),seq(0,120,30),matrix(Gb,nrow=5,byrow=TRUE)   ,main="Gb (Original Gb matrix with Kx)")
```

```
       # PCA of the subject-specific level (Gb: Gbetween) and of the subject- and visit-specific level (Gw: Gwithin):

eigen(Gb)
eigen(Gw)

       # Decide the number of components that are kept at level 1 and 2. A general rule is to stop at the component where
       # the cumulative percentage of variance explained is greater than 90% and the variance explained by any single component
       # after is less than 1/N. The number of components are also no less than the pre-determined minimum values for K1 (1) or K2 (1).

Gbpst                               <- eigen(Gb)$values/sum(eigen(Gb)$values[1:4])
Gwpst                               <- eigen(Gw)$values/sum(eigen(Gw)$values)
K1                                  <- max( which(cumsum(Gbpst) < 0.9 | Gbpst > 1/N ) + 1, 1 ) # K1 = 2
K2                                  <- max( which(cumsum(Gwpst) < 0.9 | Gwpst > 1/N ) + 1, 1 ) # K2 = 3

       # Obtain the restricted number of level 1 and 2 eigenfunctions for Gw and  Gb (some are flipped due to the physiological interpretation)

dim.space_b                         <- 2    # level 1 (subject-specific)
psi_1                               <- cbind(-eigen(Gb)$vectors[,1],-eigen(Gb)$vectors[,2])

dim.space_w                         <-  3   # level 2 (subject/visit-specific)
psi_2                               <- cbind(-eigen(Gw)$vectors[,1],-eigen(Gw)$vectors[,2],eigen(Gw)$vectors[,3])

       # Plots of the FPC harmonics
X11()
par(mfrow=c(1,2))
plot(seq(0,120,30),psi_1[,1],type="l",ylim=c(-0.7,0.8),col="dark blue"    ,
              main=paste("Subj FPC",1:dim.space_b,", based on Gb, % variance:",round(Gbpst[1:dim.space_b],2)),ylab="",xlab="")
lines(seq(0,120,30),psi_1[,1],type="l",ylim=c(-1,1),lw=12,col="dark blue")
lines(seq(0,120,30),psi_1[,2],lw=6,col="blue")
X11()
plot(seq(0,120,30),psi_2[,1],type="l",ylim=c(-0.7,0.8),col="dark blue"    ,
              main=paste("Subj/visit FPC",1:dim.space_w,", based on Gw, % variance:",round(Gwpst[1:dim.space_w],2)),ylab="",xlab="")
lines(seq(0,120,30),psi_2[,1],type="l",ylim=c(-1,1),lw=12,col="dark blue")
lines(seq(0,120,30),psi_2[,2],lw=6,col="blue")
lines(seq(0,120,30),psi_2[,3],lw=2,col="light blue")

psi_subj                            <- psi_1
psi_subvis_w                        <- psi_2


       # Minimal smoothing of FPC vectors, to optain continuous FPC curves in the plots
       # (Necessary due to the small number of glucose measurements per woman)

psi.subj.smooth                     <- smooth.basis(c(0,30,60,90,120),psi_subj,fdParobj.m )
psi.subvis.smooth.w                 <- smooth.basis(c(0,30,60,90,120),psi_subvis_w,fdParobj.m )

       # Plots of the (minimally smoothed) FPC harmonics
X11()
par(mfrow=c(1,2))
plot(psi.subj.smooth)                                           # Empirical basis functions, subject level
plot(psi.subvis.smooth.w)                                       # Empirical basis functions, subj/visit level
```

```
       # Plot of mean curves + or - 2*SD of FPCs
X11()
par(mfrow=c(3,3))
evb                          <- eigen(Gb)$values/sum(eigen(Gb)$values[1:4])
evgw                         <- eigen(Gw)$values/sum(eigen(Gw)$values[1:5])
for(i in 1:dim.space_b){
plot(overallmean.eval.smooth$fd      ,col="grey", ylim=c(2.8,8.5),lw=6,ylab=paste("Glucose (mmol/l)"),xlab="Time (min)",
                             main=paste("Overall mean \n Subject-specific FPC",i))
points(6*seq(0:20)-6, eval.fd(6*seq(0:20)-6,
       smooth.basis(c(0,30,60,90,120),(overallmean+2*sqrt(eigen(Gb)$values[i])   *(psi_subj[,i])),fdParobj.m)$fd ), pch="+")
points(6*seq(0:20)-6, eval.fd(6*seq(0:20)-6,
       smooth.basis(c(0,30,60,90,120),(overallmean-2*sqrt(eigen(Gb)$values[i])   *(psi_subj[,i])),fdParobj.m)$fd ), pch="-")
}
plot.new()
for(i in 1:dim.space_w){
plot(visitspecificmean1.eval.smooth$fd      ,col="grey", ylim=c(2.8,8.5),lw=6,ylab=paste("Glucose (mmol/l)"), xlab="Time (min)",
                             main=paste("Mean, wks 14-16 \n Subj- and visit-specific FPC",i))
points(6*seq(0:20)-6, eval.fd(6*seq(0:20)-6,
       smooth.basis(c(0,30,60,90,120),(visitspecificmean1+2*sqrt(eigen(Gw)$values[i])   *(psi_subvis_w[,i])),fdParobj.m)$fd ), pch="+")
points(6*seq(0:20)-6, eval.fd(6*seq(0:20)-6,
       smooth.basis(c(0,30,60,90,120),(visitspecificmean1-2*sqrt(eigen(Gw)$values[i])   *(psi_subvis_w[,i])),fdParobj.m)$fd ), pch="-")
}

for(i in 1:dim.space_w){
plot(visitspecificmean3.eval.smooth$fd      ,col="grey", ylim=c(2.8,8.5),lw=6,ylab=paste("Glucose (mmol/l)"),xlab="Time (min)",
                             main=paste("Mean, wks 30-32 \n Subj- and visit-specific FPC",i))
points(6*seq(0:20)-6, eval.fd(6*seq(0:20)-6,
       smooth.basis(c(0,30,60,90,120),(visitspecificmean3+2*sqrt(eigen(Gw)$values[i])   *(psi_subvis_w[,i])),fdParobj.m)$fd ), pch="+")
points(6*seq(0:20)-6, eval.fd(6*seq(0:20)-6,
       smooth.basis(c(0,30,60,90,120),(visitspecificmean3-2*sqrt(eigen(Gw)$values[i])   *(psi_subvis_w[,i])),fdParobj.m)$fd ), pch="-")
}
```

```
            ### WINBUGS

dim.space_b                   <- 2    # Antall egenfunksjoner, level 1 subject-specific
dim.space_w                   <- 3    # Antall egenfunksjoner, level 2 subject/visit-specific

psi_subj                      <- psi_subj
psi_subvis                    <- psi_subvis_w

            # The matrices W_1 and W_2 contain centered data from visits 1 and 2, respectively..

W_1                           <- as.matrix(g1.demeaned) # dim 884,5
W_2                           <- as.matrix(g3.demeaned) # dim 884,5

            # Define the data, which contains the dimension of the level 1 space, dim.space_b, the dimension of the level 2 space, dim.space_w,
            # the level 1 and 2 eigenfunctions, psi_1 and psi_2, the data matrices for visit 1 and 2,
            # the number of subjects, N_subj, the maximum number of observations per subject, N_obs

data                          <- list("dim.space_b","dim.space_w","psi_subj","psi_subvis","W_1","W_2","N_subj","N_obs")

            # Define the program file (see below)

program.file.name             <- "M:/mfpca_n884_2fpcLevel1_3fpcLevel2.txt"        # See code below


            # Define the initial values

inits.W_1                     <- matrix(rep(NA,N_subj*N_obs),ncol=N_obs)
inits.W_1[is.na(W_1)]         <- mean(mean(W_1,na.rm=TRUE))

inits.W_2                     <- matrix(rep(NA,N_subj*N_obs),ncol=N_obs)
inits.W_2[is.na(W_2)]         <- mean(mean(W_2,na.rm=TRUE))

inits.ll_b                    <- rep(0.01,dim.space_b)
inits.ll_w                    <- rep(0.01,dim.space_w)

inits                         <-function(){list(xi=matrix(rep(0,N_subj*dim.space_b),ncol=dim.space_b),
                                       zi=array(rep(0,N_subj*dim.space_w*2),c(N_subj,dim.space_w,2)),
                                       taueps=0.01,ll_b=inits.ll_b,ll_w=inits.ll_w,W_1=inits.W_1,W_2=inits.W_2)}

            # Define the parameters to be monitored

parameters=list("lambda_b","xi[1:884,]","zi[1:884,,]")
#parameters=list("lambda_b","xi[1:11,]","zi[1:11,,]") (see comment on monitoring/convergence below)

library(R2WinBUGS) # May need to install it first: install.packages("R2WinBUGS")
```

```
          # Define the thinning, iteration and burn-in numbers for the MCMC simulation

set.seed(2708)                # choose a number
n.thin         <- 100         # this number is based on test-runs with close monitoring of a selected sub-sample of some of the parameters
n.iter         <- 105000      # chosen on basis of the thinning, burn-in
n.burnin       <- 5000        # convergence begins to stabilize around 2500, some structure in some curves until 3500, chooses 5000 to be sure.

ptm <- proc.time()

Bayes.fit      <- bugs(data, inits, parameters, model.file = program.file.name,
                           n.chains = 1, n.iter = n.iter, n.burnin = n.burnin,
                           n.thin = n.thin, debug = FALSE, DIC = FALSE, digits = 5,
                           codaPkg = FALSE,
                           bugs.directory = "D:/winbugs14/WinBUGS14/")

proc.time() - ptm

autocorr.plot(as.mcmc.list(Bayes.fit), lag.max=50, auto.layout = TRUE)    # only works if few parameters are monitored,
                                                                          # e.g. "lambda_b","xi[1:11,]","zi[1:11,,]"
print(Bayes.fit)
plot(Bayes.fit)
head(Bayes.fit)
attach.bugs(Bayes.fit)

scores.l1.subj                <- colMeans(xi)               # Subject-specific FPC scores
scores.l2.subj.v1             <- colMeans(zi)[,,1]          # Subject- and visit-specific FPC scores, visit 1
scores.l2.subj.v3             <- colMeans(zi)[,,2]          # Subject- and visit-specific FPC scores, visit 3

       # Correlation table

round(cor(cbind(eval.g1,eval.g3,spssdata.term.complete.ogtt$auc1,spssdata.term.complete.ogtt$auc3,
                                          scores.l1.subj,scores.l2.subj.v1,scores.l2.subj.v3) ),2)

       # Save FPC scores

scores.mfpca                  <- cbind(spssdata.term.complete.ogtt$id,scores.l1.subj,scores.l2.subj.v1,scores.l2.subj.v3)
write.table(scores.mfpca, file="M:/Art5longitudinalFDA/Bayes_FDA/R2WinBUGS/scores.mfpca.may2013.csv")

       # Read saved FPC scores from file
scores.mfpca                  <- read.table("M:/Art5longitudinalFDA/Bayes_FDA/R2WinBUGS/scores.mfpca.may2013.csv")
```

```
       ### FIGURE 2

X11()   #       The 2*884 smoothed glucose curves
par(mar=c(1,1,1,1))
plot(g1.smooth,lty=1,col="black",lw=2,ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
plot(g3.smooth,lty=1,col="black",lw=2,add=TRUE)
abline(h=0,col="grey",lw=5)

X11()   #       Overall mean
par(mar=c(1,1,1,1))
plot(overallmean.eval.smooth,lty=1,lw=5,col="black",ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
abline(h=0,col="grey",lw=5)

X11()   #       eta
par(mar=c(1,1,1,1))
plot(visitspecificmean1.eval.smooth$fd-overallmean.eval.smooth$fd,lty=1,lw=5,col="black",ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
plot(visitspecificmean3.eval.smooth$fd-overallmean.eval.smooth$fd,lty=1,lw=5,col="black",xaxt='n', yaxt='n',ann=FALSE,add=TRUE)
abline(h=0,col="grey",lw=5)


X11()   #       The estimated X-curves
par(mar=c(1,1,1,1))
plot(scores.mfpca[884,2]*psi.subj.smooth$fd[1]+scores.mfpca[884,3]*psi.subj.smooth$fd[2],
               lty=1,lw=2,col="black",ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
for(i in 1:884){
plot(scores.mfpca[i,2]*psi.subj.smooth$fd[1]+scores.mfpca[i,3]*psi.subj.smooth$fd[2],lty=1,lw=2,col="black",xaxt='n',yaxt='n',ann=FALSE,add=TRUE)}
abline(h=0,col="grey",lw=5)

X11()   #       The estimated U-curves
par(mar=c(1,1,1,1))
plot(scores.mfpca[828,4]*psi.subvis.smooth.w$fd[1]+scores.mfpca[828,5]*psi.subvis.smooth.w$fd[2]+scores.mfpca[828,6]*psi.subvis.smooth.w$fd[3],
               lty=1,lw=2,col="black",ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
plot(scores.mfpca[828,7]*psi.subvis.smooth.w$fd[1]+scores.mfpca[828,8]*psi.subvis.smooth.w$fd[2]+scores.mfpca[828,9]*psi.subvis.smooth.w$fd[3],
               lty=1,lw=2,col="black", xaxt='n', yaxt='n',ann=FALSE,add=TRUE)
for(i in 1:884){
plot(scores.mfpca[i,4]*psi.subvis.smooth.w$fd[1]+scores.mfpca[i,5]*psi.subvis.smooth.w$fd[2]+scores.mfpca[i,6]*psi.subvis.smooth.w$fd[3],
               lty=1,lw=2,col="black", xaxt='n', yaxt='n',ann=FALSE,add=TRUE)
plot(scores.mfpca[i,7]*psi.subvis.smooth.w$fd[1]+scores.mfpca[i,8]*psi.subvis.smooth.w$fd[2]+scores.mfpca[i,9]*psi.subvis.smooth.w$fd[3],
               lty=1,lw=2,col="black", xaxt='n', yaxt='n',ann=FALSE,add=TRUE)
}
abline(h=0,col="grey",lw=5)

X11()   #       Visit-specific means
par(mar=c(1,1,1,1))
plot(visitspecificmean1.eval.smooth,lty=1,lw=5,col="black",ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
plot(visitspecificmean3.eval.smooth,lty=1,lw=5,col="black", xaxt='n', yaxt='n',ann=FALSE,add=TRUE)
abline(h=0,col="grey",lw=5)

X11()   #       B-splines-smoothed curves for woman no 828
par(mar=c(1,1,1,1))
plot(g1.smooth$fd[828],lty=1,lw=5,col="black",ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
plot(g3.smooth$fd[828],lty=1,lw=5,col="black",add=TRUE)
```

```
abline(h=0,col="grey",lw=5)

X11()   #       Estimated X-curve for woman no 828
par(mar=c(1,1,1,1))
plot(scores.mfpca[828,2]*psi.subj.smooth$fd[1]+scores.mfpca[828,3]*psi.subj.smooth$fd[2],
               lty=1,lw=5,col="black",ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
abline(h=0,col="grey",lw=5)

X11()   #       mu(t) + eta(t) + Xhat(t) for woman no 828
par(mar=c(1,1,1,1))
plot(visitspecificmean1.eval.smooth$fd+scores.mfpca[828,2]*psi.subj.smooth$fd[1]+scores.mfpca[828,3]*psi.subj.smooth$fd[2],
               lty=1,lw=5,col="black",ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
plot(visitspecificmean3.eval.smooth$fd+scores.mfpca[828,2]*psi.subj.smooth$fd[1]+scores.mfpca[828,3]*psi.subj.smooth$fd[2],
               lty=1,lw=5,col="black", xaxt='n', yaxt='n',ann=FALSE,add=TRUE)
abline(h=0,col="grey",lw=5)

X11()   #       Estimated U-curves for woman no 828
par(mar=c(1,1,1,1))
plot(scores.mfpca[828,4]*psi.subvis.smooth.w$fd[1]+scores.mfpca[828,5]*psi.subvis.smooth.w$fd[2]+scores.mfpca[828,6]*psi.subvis.smooth.w$fd[3],
               lty=1,lw=5,col="black",ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
plot(scores.mfpca[828,7]*psi.subvis.smooth.w$fd[1]+scores.mfpca[828,8]*psi.subvis.smooth.w$fd[2]+scores.mfpca[828,9]*psi.subvis.smooth.w$fd[3],
               lty=1,lw=5,col="black", xaxt='n', yaxt='n',ann=FALSE,add=TRUE)
abline(h=0,col="grey",lw=5)

X11()   #       mu(t) + eta(t) + Xhat(t) + Uhat(t) for woman no 828
par(mar=c(1,1,1,1))
plot(visitspecificmean1.eval.smooth$fd+scores.mfpca[828,2]*psi.subj.smooth$fd[1]+scores.mfpca[828,3]*psi.subj.smooth$fd[2]+

       scores.mfpca[828,4]*psi.subvis.smooth.w$fd[1]+scores.mfpca[828,5]*psi.subvis.smooth.w$fd[2]+scores.mfpca[828,6]*psi.subvis.smooth.w$fd[3],
               lty=1,lw=5,col="black",ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
plot(visitspecificmean3.eval.smooth$fd+scores.mfpca[828,2]*psi.subj.smooth$fd[1]+scores.mfpca[828,3]*psi.subj.smooth$fd[2]+

       scores.mfpca[828,7]*psi.subvis.smooth.w$fd[1]+scores.mfpca[828,8]*psi.subvis.smooth.w$fd[2]+scores.mfpca[828,9]*psi.subvis.smooth.w$fd[3],
               lty=1,lw=5,col="black", xaxt='n', yaxt='n',ann=FALSE,add=TRUE)
abline(h=0,col="grey",lw=5)

X11()   #       B-splines-smoothed curves for woman no 828 and mu(t) + eta(t) + Xhat(t) + Uhat(t) for woman no 828 in the same plot
par(mar=c(1,1,1,1))
plot(g1.smooth$fd[828],lty=1,lw=5,col="black",ylim=c(-2,12.5),xaxt='n', yaxt='n',ann=FALSE)
plot(g3.smooth$fd[828],lty=1,lw=5,col="black",add=TRUE)
plot(visitspecificmean1.eval.smooth$fd+scores.mfpca[828,2]*psi.subj.smooth$fd[1]+scores.mfpca[828,3]*psi.subj.smooth$fd[2]+

       scores.mfpca[828,4]*psi.subvis.smooth.w$fd[1]+scores.mfpca[828,5]*psi.subvis.smooth.w$fd[2]+scores.mfpca[828,6]*psi.subvis.smooth.w$fd[3],
               lty=2,lw=5,col="black",xaxt='n', yaxt='n',ann=FALSE,add=TRUE)
plot(visitspecificmean3.eval.smooth$fd+scores.mfpca[828,2]*psi.subj.smooth$fd[1]+scores.mfpca[828,3]*psi.subj.smooth$fd[2]+

       scores.mfpca[828,7]*psi.subvis.smooth.w$fd[1]+scores.mfpca[828,8]*psi.subvis.smooth.w$fd[2]+scores.mfpca[828,9]*psi.subvis.smooth.w$fd[3],
               lty=2,lw=5,col="black", xaxt='n', yaxt='n',ann=FALSE,add=TRUE)
abline(h=0,col="grey",lw=5)
```

The program file, "M:/mfpca_n884_2fpcLevel1_3fpcLevel2.txt":

```
model
{#Start model
       for (i in 1:N_subj)
    {for (t in 1:N_obs)
        {W_1[i,t]~dnorm(m_1[i,t],taueps)
         W_2[i,t]~dnorm(m_2[i,t],taueps)

         m_1[i,t]<-X[i,t]+U_1[i,t]
         m_2[i,t]<-X[i,t]+U_2[i,t]

         X[i,t]       <-xi[i,1]*psi_subj[t,1]+xi[i,2]*psi_subj[t,2]

         U_1[i,t]     <-zi[i,1,1]*psi_subvis[t,1]+zi[i,2,1]*psi_subvis[t,2]+zi[i,3,1]*psi_subvis[t,3]

         U_2[i,t]     <-zi[i,1,2]*psi_subvis[t,1]+zi[i,2,2]*psi_subvis[t,2]+zi[i,3,2]*psi_subvis[t,3]
        }

     for (k in 1:dim.space_b)
        {xi[i,k]~dnorm(0,ll_b[k])}

     for (l in 1:dim.space_w)
        {zi[i,1,1]~dnorm(0,ll_w[l])
         zi[i,1,2]~dnorm(0,ll_w[l])}
        }#

for (k in 1:dim.space_b)
        {ll_b[k]~dgamma(1.0E-3,1.0E-3)
         lambda_b[k]<-1/ll_b[k]}

for (l in 1:dim.space_w)
        {ll_w[l]~dgamma(1.0E-3,1.0E-3)
         lambda_w[l]<-1/ll_w[l]}

taueps~dgamma(1.0E-3,1.0E-3)
sigma_sq_eps<-1/taueps
}#End model
```