Longitudinal vs. cross-sectional design

One of the most heated debates is related to the study design, in particular to the option about measuring each involved foetus once (cross-sectional study) or more times (longitudinal study).

Serial measurements in order to assess fetal growth were first introduced by Willocks in 1962 and the first longitudinal chart for biparietal diameter was presented by Campbell and Newman in 1971. However, many obstetricians use cross-sectional charts for their serial fetal examinations, even though cross-sectional studies are not designed to evaluate growth. Royston points out that when using size-charts instead of growth charts to determine growth, the obstetrician may be misled. In contrast to cross-sectional data, longitudinal data have a hierarchical structure based on two levels of variation: within foetuses between gestational ages (level 1), and variation between foetuses (level 2).

In this paper we will ignore the serial nature of the data, calculating the percentiles as if the data were derived from separate individuals.

*This means that I’m trying to obtain the reference centile equations by exploiting all available data and by considering each ultrasound visit as it were atomic even if the same pregnant woman will be visited for more than one time.*

*Which might be the consequences of such choice?*

Does this mean that you want to split the data for example looking at one factor from all the data and work out centile equations on all factors seperatly and then looking at the over all growth without knowing which factors affect the other??

This might be good for looking at purely growth patterns but not so good at determining consequences of the growth patten. By this I mean, you will be able to see that your baby is big, small or normal but it won't tell you the consequences.

Sample “numerosity”

The “*numerosity*” of the sample is questionable: the determination of the suitable sample size for fetal growth studies is a quite complex matter. Discussions about it are in the most typical introductory biostatistics texts, but the key factor is about the role of sample size in statistical analysis: it’s significant primarily because of its effect on statistical power.

Two common mistakes in sample size determination of observational studies are:

* the goodness of the results obtained from a sample depends solely on the number of included individuals and not from the way in which they were selected;
* a suitable sample size should be proportional to the size of the study population.

Obviously, the larger the sample, and more accurate and reliable will be results, on condition that the sample has been selected with a proper method, but the laws of statistics show that the conventional wisdom mentioned above are completely false.

*What can you tell me about it?*

We need to link this to your frame-work proposal. Saying something along the lines of “we will not need to take random samples from a population, all the data will be stored directly into this database” the only problem we'll have is waiting until the data collected is big enough to represent a population. We won't have to worry about what way it's selected, as it will automatically be collected at every pregnancy ??

Sample size can also be linked to a confidence level, using formulas for confidence that take into account the sample size. Now the quality of the sample is a completely different matter, this is another opportunity to mention your frame-work, mentioning that it will help eliminate noisey and corrupted data.

In the context of fetal growth, knowing that every year in the world there are about 160 millions of newborns, the determination of an appropriate sample size is not so straightforward.

*Which should be our “right size”?*

Following from what I mentioned earlier, sample size could improve your confidence level. But you have to remember that you shouldn't get out caught in a trap, don't worry about choosing a sample size, be worried about whether your sample size is big enough. Also you can to get growth for different reasons and mentioning the whole world population is a big useless, like the example you gave about Chinese babies in Italy, similarly an Italian mother won't want to worry about growth rates in Norway right, not even if every fetus has been measured since the 60s there.

Linear regression

An important and essential assumption for the mathematical modelling of the phenomenon is about the samples’ distribution, that is a Gaussian distribution with a mean and a standard deviation by which is possible to identify that a certain number of samples satisfy some average value of a certain biometric parameter. This assumption allows defining the centile curve by using the well-known formula: *Centile = mean + K \* SD*

where K is the corresponding centile of the standard Gaussian distribution.

In order to obtain the interpolating curve to the experimental data, we have adopted the Least Mean Squares (LMS) technique, which is based on the minimization of the sum of squares of the deviations of each measure by the approximating curve, using a polynomial of any order in order to make the "best fit". Being the curves modelled by polynomials, we talk about multiple linear regression.

*By considering the* ***mean****:*

*How can I find the best fit? I know that there are almost 3 different approaches:*

1. *matrioska model: we start from a third order polynomial; if the cubic coefficient is not significantly different from zero (approximately if it is less than twice its SD, right?)  second order polynomial will be fitted with the same assessment made of the quadratic coefficient. The process should be repeated until no further removal of terms is possible.*
2. *Evaluation of the index of determination R2 which accepts values between 0 and 1 and that approaches much more to 1 as the interpolating curve approaches to real data. Is it sufficient only the coefficient of determination for evaluating the quality of a given model with respect to another one?*

*Yes, the coefficient of determination is a percentage stating how much of the regression equation passes through the actual points, (have to remember than the regression is just an approximation). So we can calculate the coefficient of determination for different lines against the same data and see which percentage is higher. This doesn't help you find a regression though, just to see how good it is.*

1. *Evaluation of coefficient significance (pvalue, sterror, tvalue). Which intervals of significance should I consider, what kind of thresholds? For example: pvalue < 0.05  significant regression?*

*P-values are only useful for showing that the model isn't rubbish, but they are NOT sufficient to tell us is the data is actually good, do you understand what I mean? Like having a low P-value is good, but not enough to show that our model is correct. ST-error is used to determine whether a sample population mean is accurate (compared to the actual size), this can be used to prove that your population is big enough for meaningful results. T-values are used for comparing regressions and also used to determine P-values. I'm not 100% on what thresholds to have though, I think they are very problem specific so we would need to check.*

*By considering the* ***standard deviation****:*

*In order to estimate the SD we start from the scaled and absolute residuals (the absolute value of the difference between the observed values and those predicted by the regression equation, multiplied then for a corrective constant equal to √(π/2) ).*

* *if residuals do not show a trend with gestational age (x axis)  SD estimated as the SD of the original residuals (observed value – predicted value)*
* *if there is a trend  polynomial linear regression (which is almost never greater than the second order)*

*Is it right?*

Goodness of model evaluation

A useful tool in assessing model fit is the adoption of Z-scores (also known as SD scores), which are defined as:

(2)

where and are, respectively, the mean and SD given by the model for the GA at which the observation is made. Hence Z-scores represent the observed values expressed on a standard normal scale (with a mean of 0 and SD of 1), with the mean and SD adjusted for GA.

*For the evaluation of the model goodness:*

* *z scores are plotted against gestational age. They should be randomly scattered about zero at all gestational ages (typically between -1,645 and +1,645).*
  + *If there exist a pattern  the mean curve may require modification*

*What does it mean? Do I have to decrease the polynomial order?*

*I'm not sure what you are trying to do here...*

* *the QQ plot is considered in order to verify that z scores have a normal distribution.*

*Is it equivalent to perform, more formally, the Shapiro-Wilk and Shpiro Francia test able to check the model accuracy?*

*Q-Q plot are just visualisations for comparing distributions. We can do a Q-Q on our data vs a normal theoretical data to see if it follows a normal distribution. I'm not sure how we can check the model accuracy with these tests though. Just because we can say that the data is normally distribution and the model is normally distributed doesn't mean that the model is correct.*

* *The appropriate proportion of observations should fall between and outside fitted centiles, for example approximately 90% of Z-scores should lie between Z = −1.645 and Z = +1.645*

*If this happens, does it mean that the model is right?*

*If this happens, it doesn't tell us whether it's right of wrong. Remember we are talking about probabilities and approximations. What we need to focus on is what we consider a good (meaning full) score. I think if we get Z=+-2 then more of less 95% of the data should be near the mean at no more than 2 SDs away from the mean. If we have Z=+-3 then 99% .*

Correct coefficient representation

*Supposing to obtain such coefficients for the Femur Length biometric parameter:*

Macintosh HD:Users:Lucia:Desktop:Schermata 11-2456990 alle 11.07.20.png

*Is it true that it’s the more correct to represent these coefficients by considering the uncertainty (given by the Standard Error) which expresses the confidence degree that we have found on the result?*

*I don't think there is much different between showing that table or the equations below. I think it's more common to show the equations below if you plan to use them or show them mathematically whereas the table above would be shown as proof of the results you got.*

*The uncertainty should be rounded as well. It is necessary to establish the number of significant digits for an adequate representation.*

*Yes, this is good practice to round mainly to make the results nicer for the reader. The rule below is fine.*

*Typically the “rule” is to round to one significant digit. [REF: John Robert Taylor. “Introduction to Error Analysis. The Study of Uncertainties in Physical Measurements”. 1986]*

*I disagree with that statement. Rounding 3.14 to one significanct digit gives you 3 Maybe they meant one significant digit from the tenths. It's a matter of preferance really. If it where me I would leave it as B = -1.24 +- 0.61*

*This would be because 0.04 is much more significant than 0.001 ie 0.04 would have to much of a big effect if we removed it.*

*Obtaining:*

*FLmean = -1,2393855381 + 0,1373541980\*GA + 0,0075373017\*GA2 – 0,0001320367\*GA3*

*Rounding:*

*FLmean = -1,2394 + 0,13735\*GA + 0,007537\*GA2 – 0,000132\*GA3*

*Should I represent coefficients in this way?*