Estimate / Guesstimate:

Is ultrasound estimated fetal weight accurate enough?

Nick Dudley, PhD

United Lincolnshire Hospitals

Disclaimer

This presentation is based on research conducted to the best of the presenter's abilities, acknowledging inherent uncertainties and limitations. The findings may be influenced by factors such as data quality, methodology, and interpretation, and do not guarantee real-world accuracy.

The content is intended for informational purposes only and does not necessarily reflect the views of the British Medical Ultrasound Society (BMUS) or its affiliates. The presenter is responsible for ensuring objectivity and disclosing any relevant financial/industry interests. BMUS and its affiliates disclaim liability for any claims arising from this educational activity.

Scope

- How it is now
  - Historical reviews
  - More recent evidence
- Where do errors come from?
  - Types
  - Sources
  - Modelling
- What can we do?
  - Methods and procedures
  - Training
  - Audit
  - AI

Questions for you

- Who uses EFW as their main growth measurement?
- Who is happy with the accuracy of EFW?
- Who uses AI / automated measurement?
- Who is happy with the accuracy of automated measurement?

Questions for you

- What is an “acceptable” accuracy?
- 20%
- 15%
- 10%
- 5%

Terminology

- Mean is an arithmetic average of measurements or errors

- Standard deviation (SD) is a measure of the spread of measurements or errors (for a Normal distribution)
  - 1 SD, ~68% of measurements fall within it
  - 2 SD, ~95% of measurements fall within it

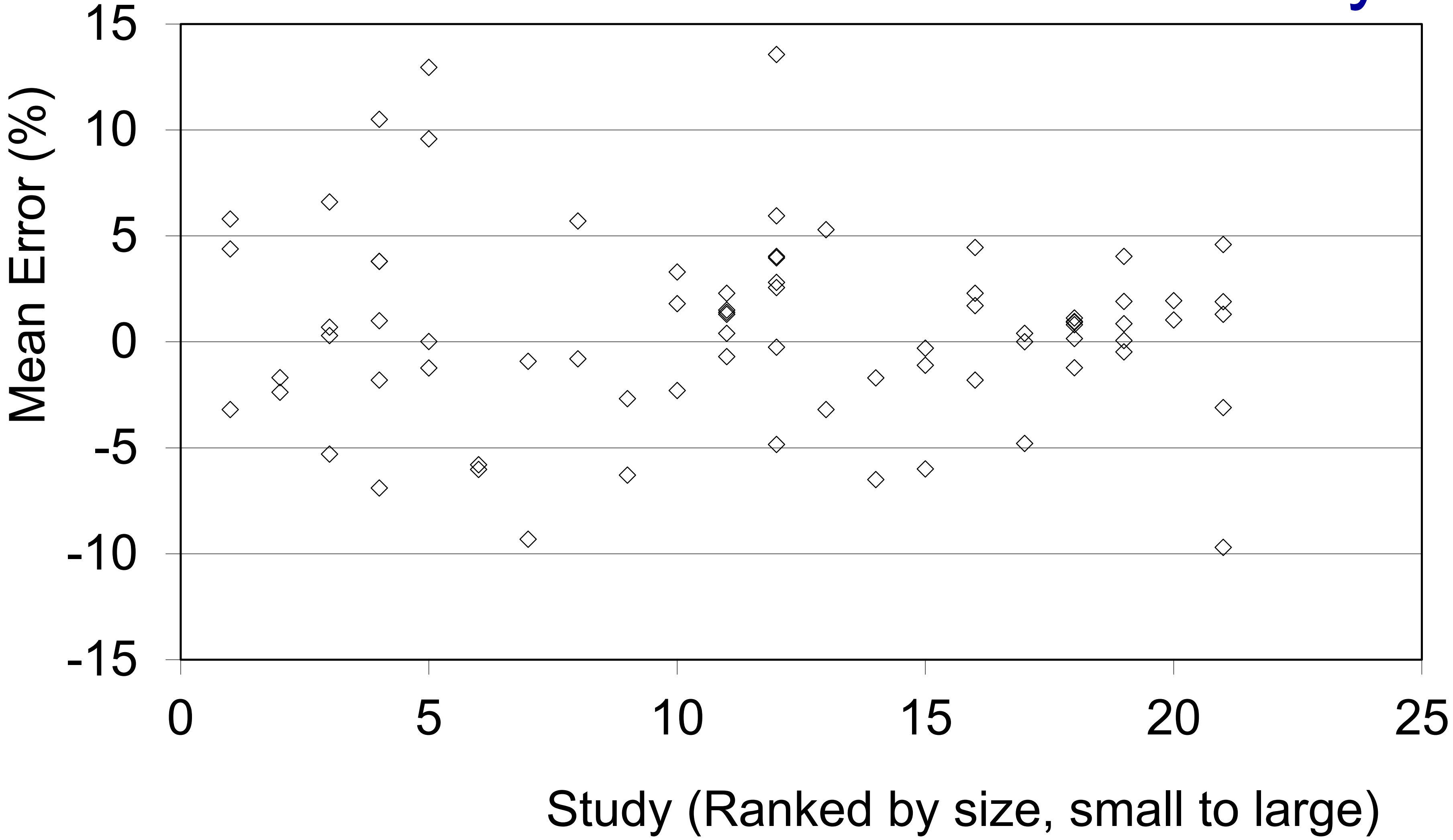
- 95% Confidence Interval, ±1.96 SD, 1 in 20 measurements fall outside it

Compound errors

- Where measurements are combined, any errors are compounded
- For example:
  - AC error 3%
  - HC error 3%
  - FL error 2%

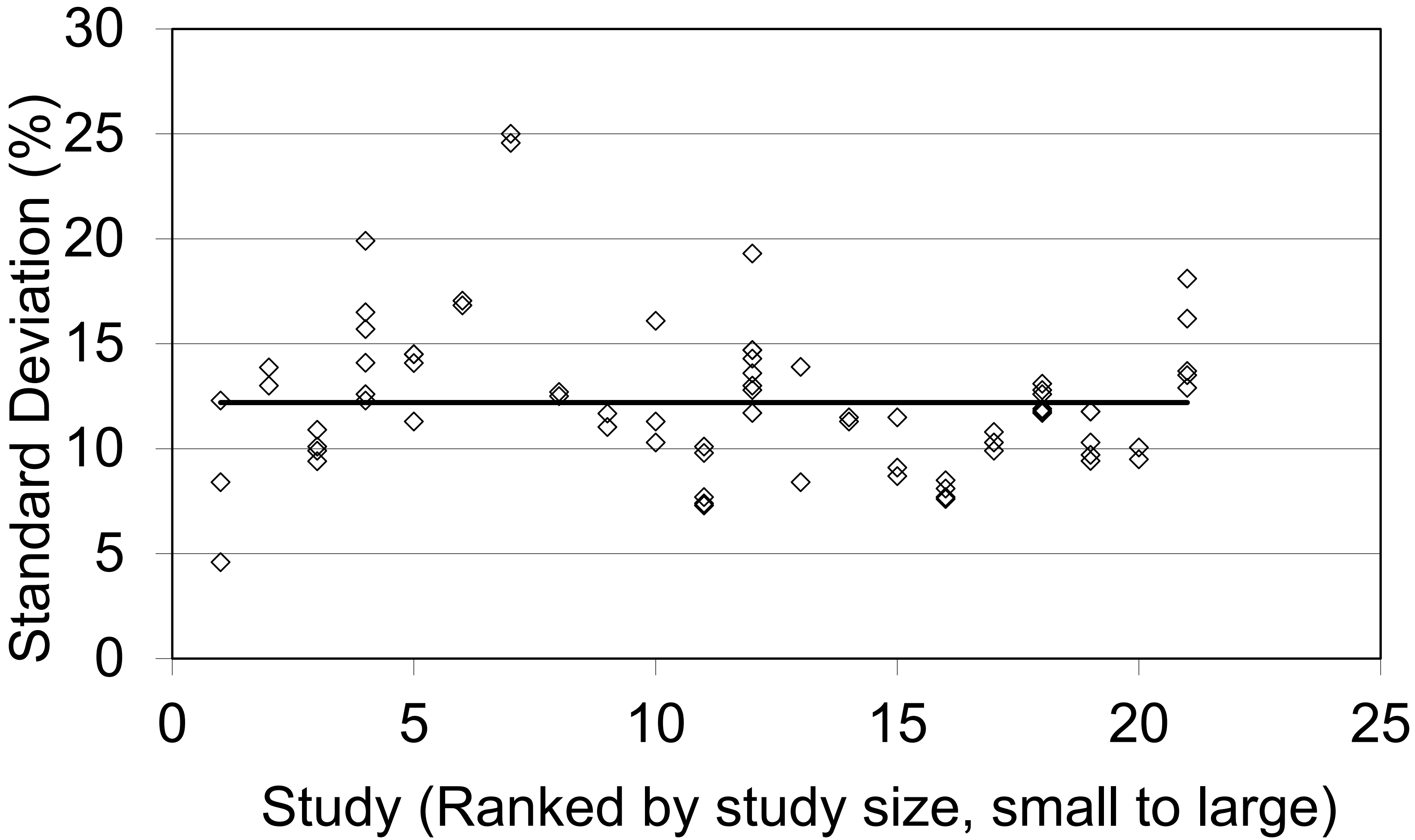
$$\text{EFW error} \approx 3\% + 3\% + 2\% = 8\%$$

Historical reviews of EFW accuracy



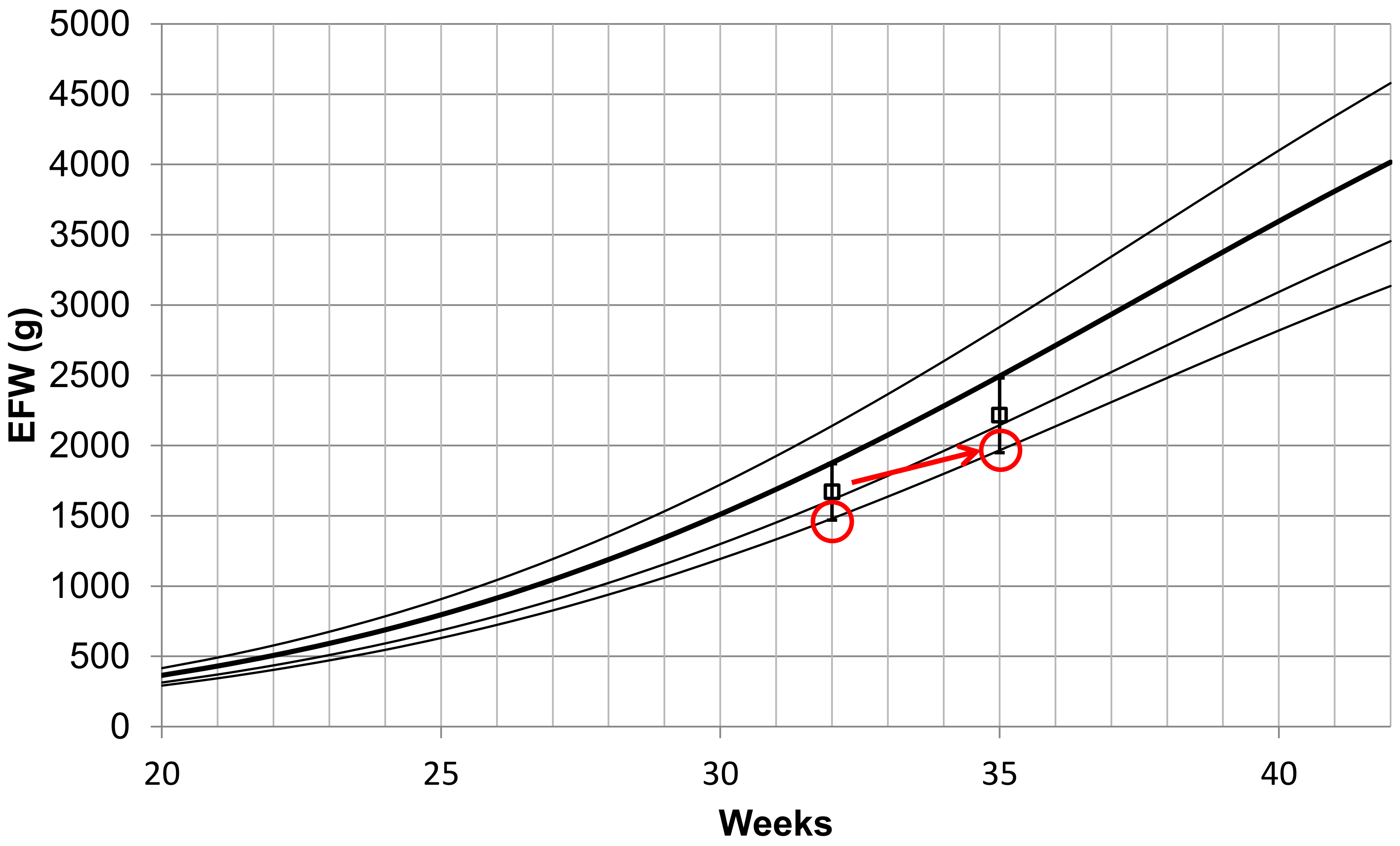
Dudley NJ. A systematic review of the ultrasound estimation of fetal weight. *Ultrasound Obstet Gynecol* 25: 80-89, 2005

Range of errors



Dudley NJ. A systematic review of the ultrasound estimation of fetal weight.
Ultrasound Obstet Gynecol 25: 80-89, 2005

Does EFW error matter?



±12% error bars

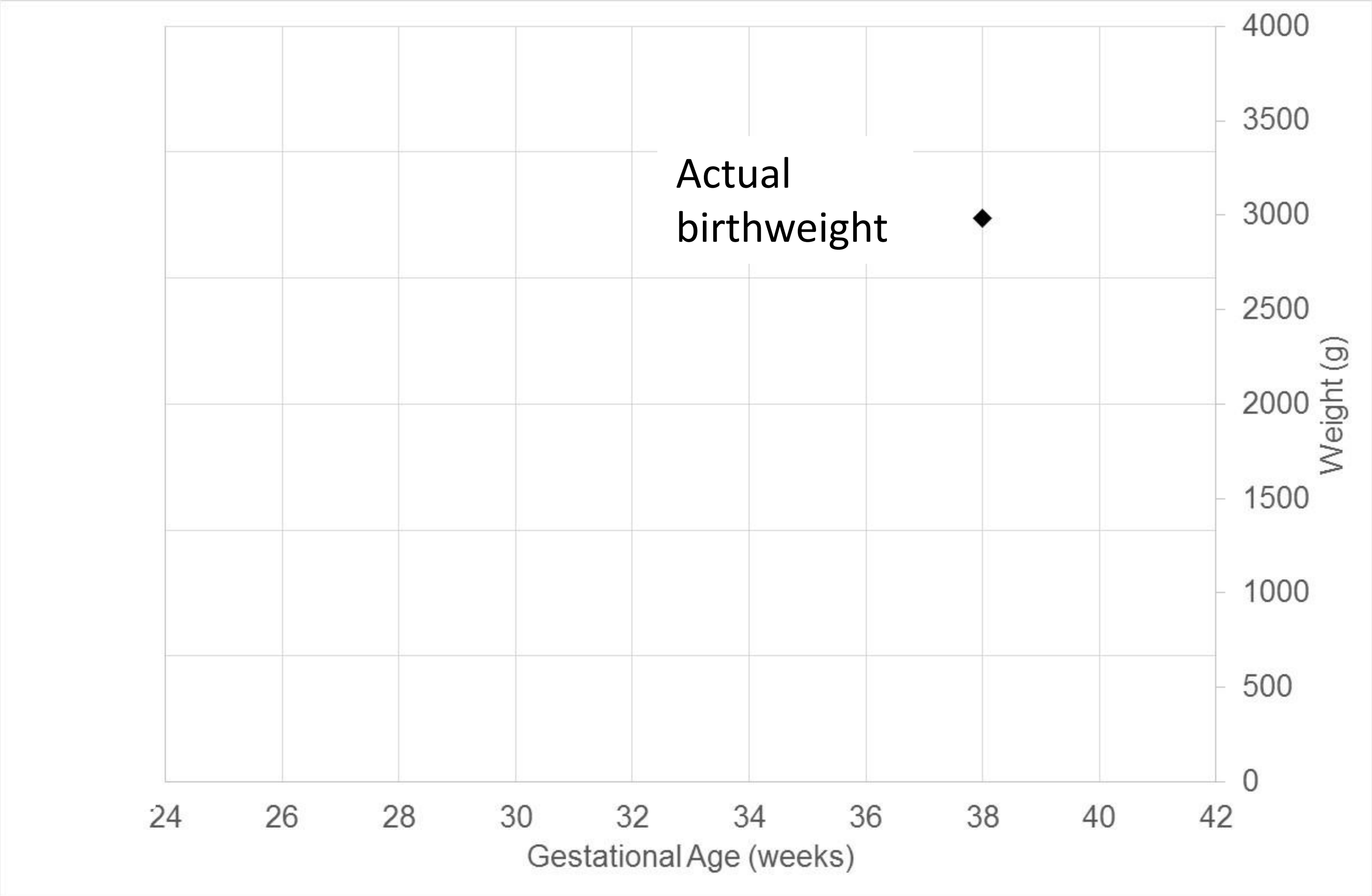
Real impact on plotting growth

- Studies of EFW accuracy need a “gold standard”.
- Usually birthweight.
- How do we know if serial scans are accurate?
- We need to compare with the growth curve for each fetus.

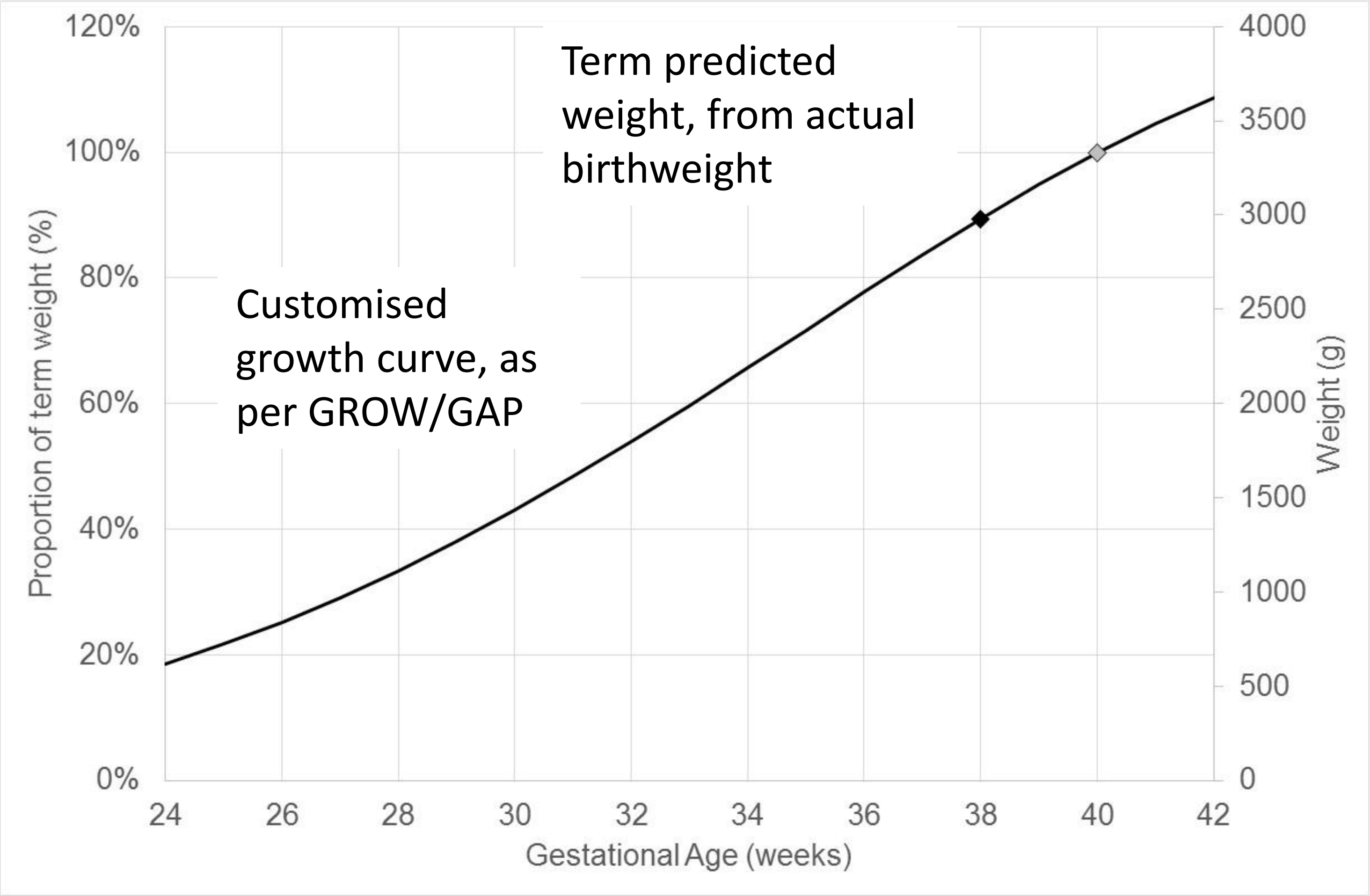
Real impact on plotting growth

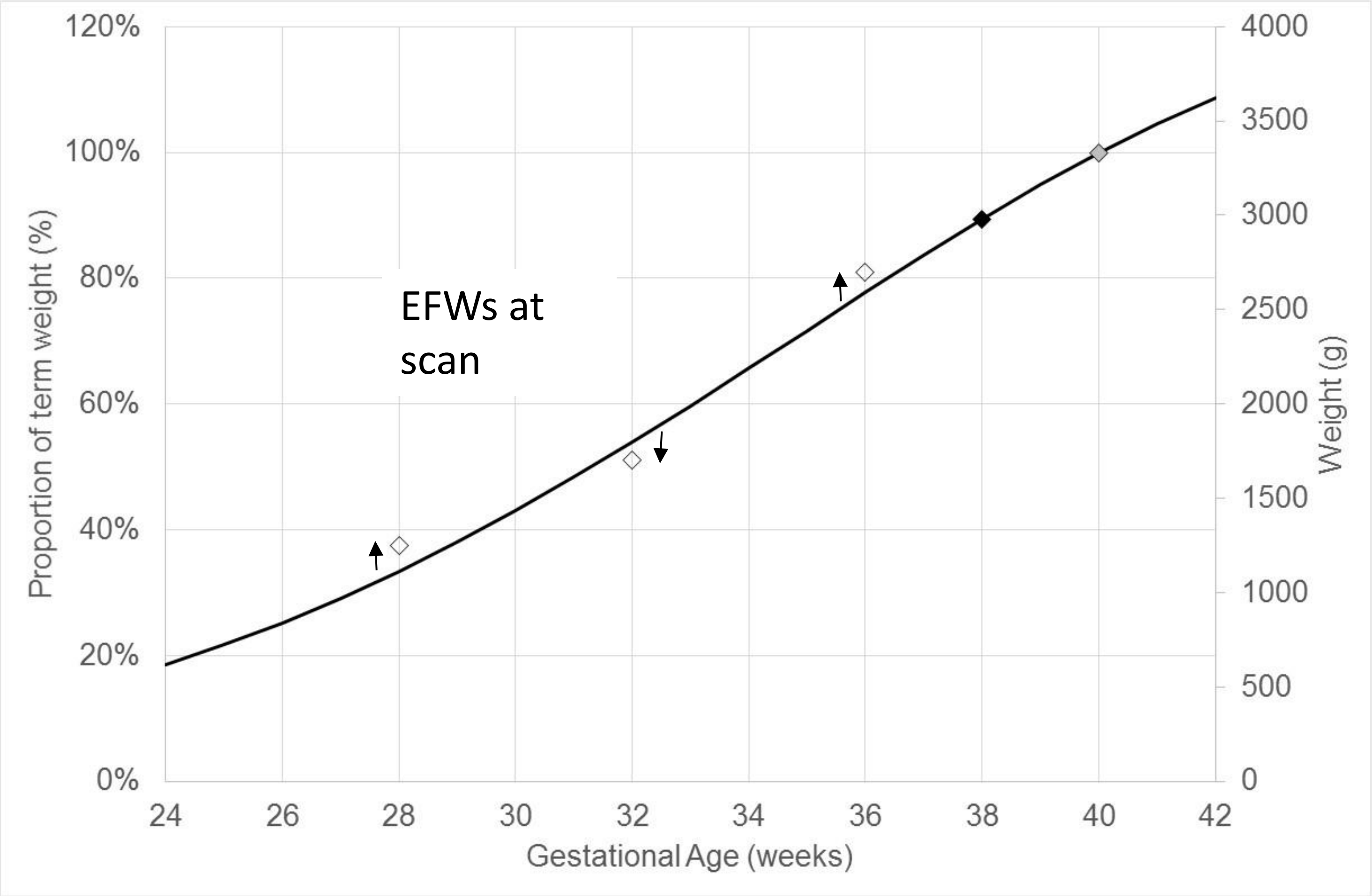
- Assumed growth curves plotted from birth weight (as GROW/GAP does using customised expected birth weight)
- Serial scans plotted on assumed growth curve
- “Errors” calculated

Dudley NJ, Jindal S, Varley H. An evaluation of systematic and random errors in ultrasound estimated fetal weight during serial ultrasound. *Ultrasound* 31: 259-265, 2023



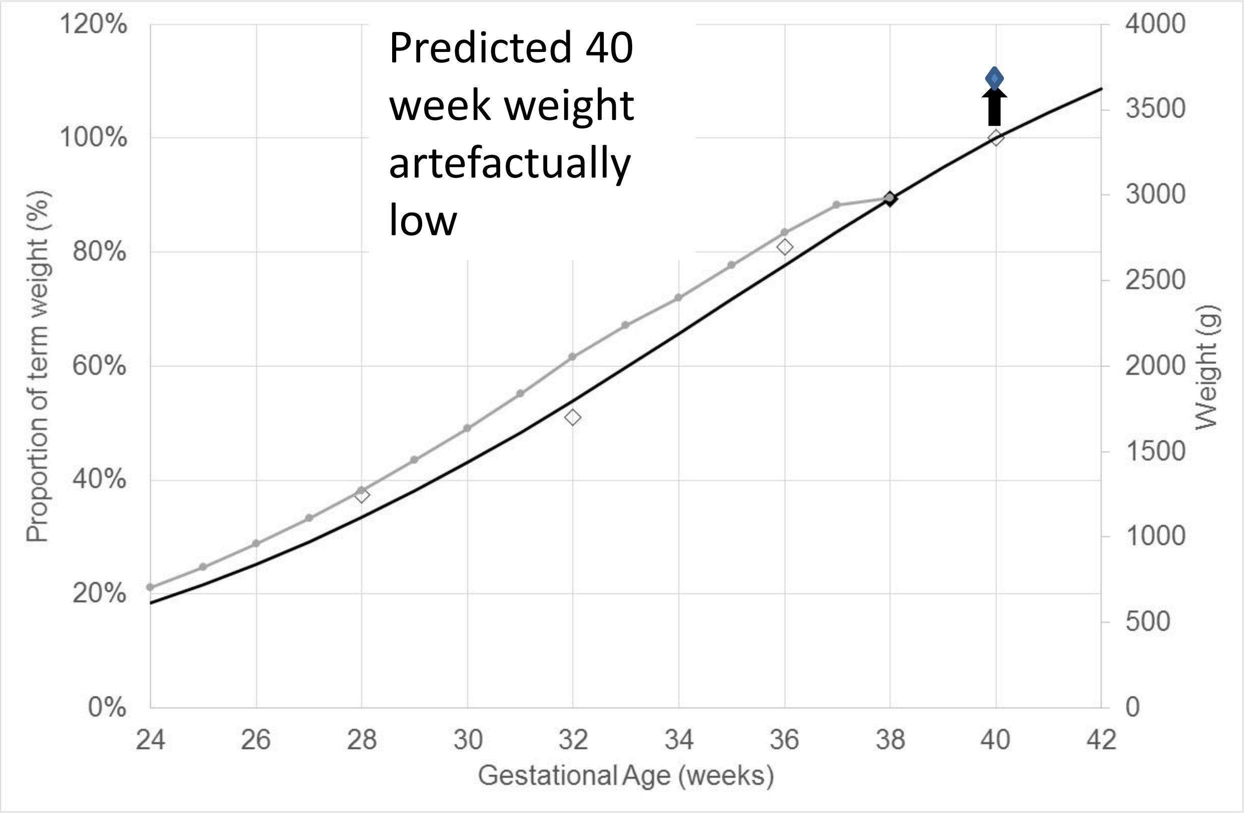
Actual
birthweight
4000
3500
3000
2500
2000
1500
1000
500
0
Weight (g)
24
26
28
30
32
34
36
38
40
42
Gestational Age (weeks)


Term predicted weight, from actual birthweight
Customised growth curve, as per GROW/GAP
120%
100%
80%
60%
40%
20%
0%
Proportion of term weight (%)
4000
3500
3000
2500
2000
1500
1000
500
0
Weight (g)
24
26
28
30
32
34
36
38
40
42
Gestational Age (weeks)



EFWs at scan
120%
100%
80%
60%
40%
20%
0%
Proportion of term weight (%)
4000
3500
3000
2500
2000
1500
1000
500
0
Weight (g)
24
26
28
30
32
34
36
38
40
42
Gestational Age (weeks)

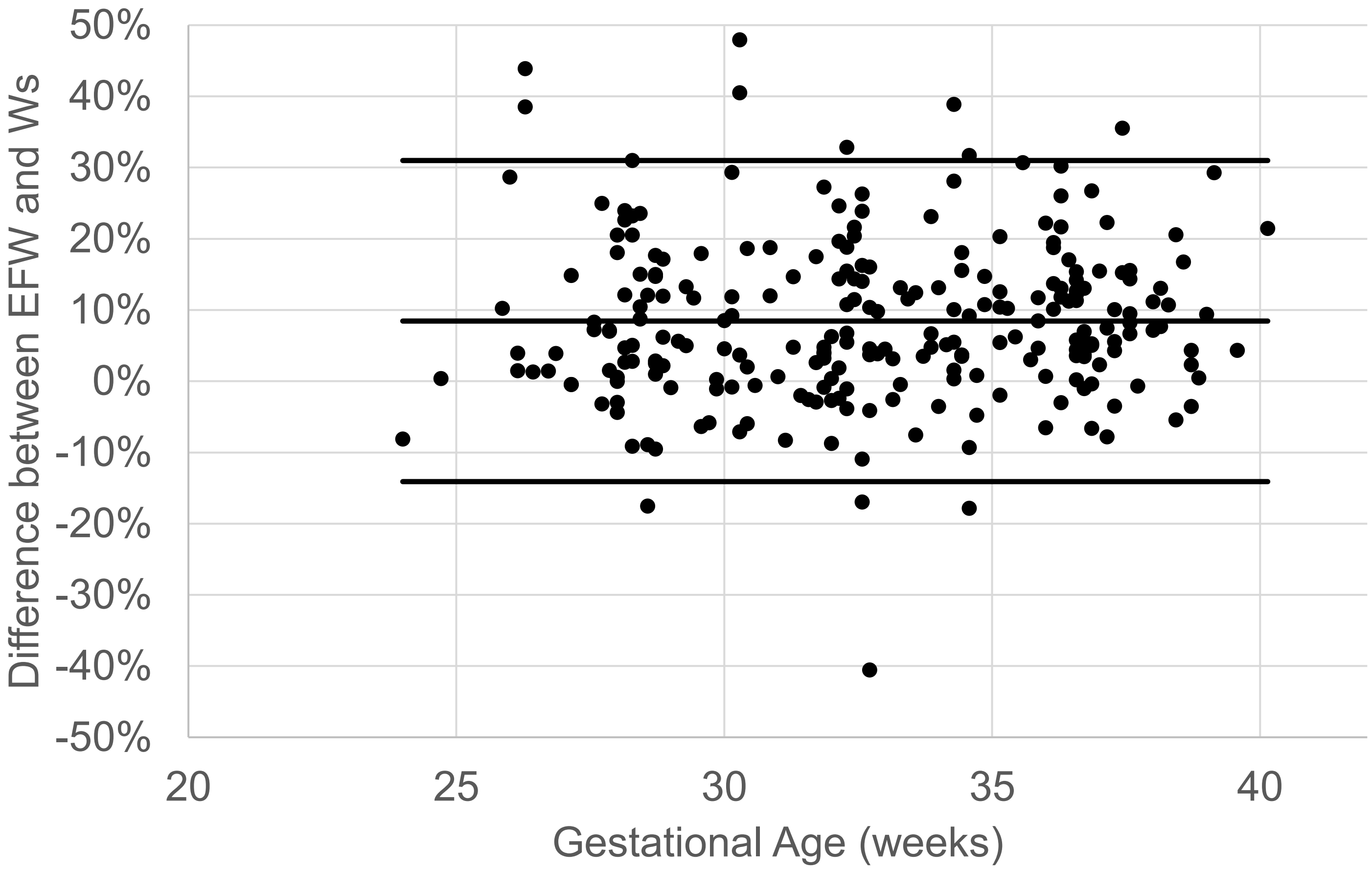
Why it doesn't work if growth falls off



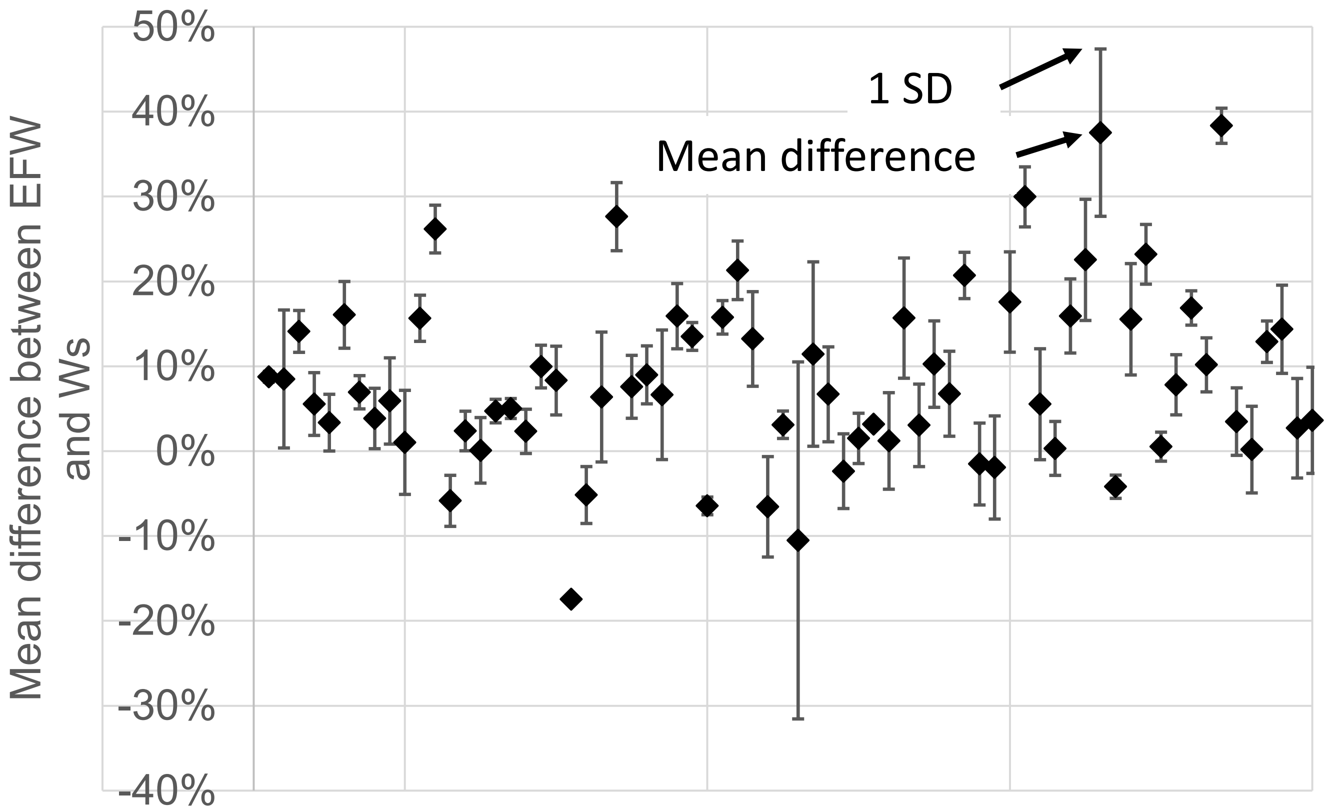
Potentially pathological growth profiles excluded from analysis

Results – Data Summary

(Ws is expected weight at scan)



Results – Individual Subjects



Results – Summary of Differences

- Range of mean differences: -17.5% to 38.3% (mean: 8.4%)
- Systematic overestimation
- Range of standard deviations: 0.4% to 21% (mean: 4.3%)
- Potentially large random errors

Impact on growth monitoring

- Errors confound growth rate monitoring and assessment of absolute size.
- If systematic error is operator independent, no problem for growth velocity,
  - e.g. if strongly related to EFW formula.
- Random error is always a confounding factor.

So …

- EFW errors can have a real impact on growth monitoring.
- Important to
  - identify potential sources of inaccuracy,
  - work to minimise errors.
- We need to have a little understanding of errors and look at each potential error source.

Errors from the literature

- 95% Confidence Intervals (1 in 20 errors larger)
- HC ±5%
- AC ±9% 5% + 9% + 11% = 25% !!
- FL ±11%

This may be “worst case”, as random errors could cancel.

But … average 95% CI from review is ±24%

Sarris I, Ioannou C, Chamberlain P et al. Intra- and interobserver variability in fetal ultrasound measurements. *Ultrasound Obstet Gynecol* 39, 266-273, 2012

Sources of EFW errors

- Measurements (AC, HC, FL)
- EFW formula
- Interpretation (charts and procedures)

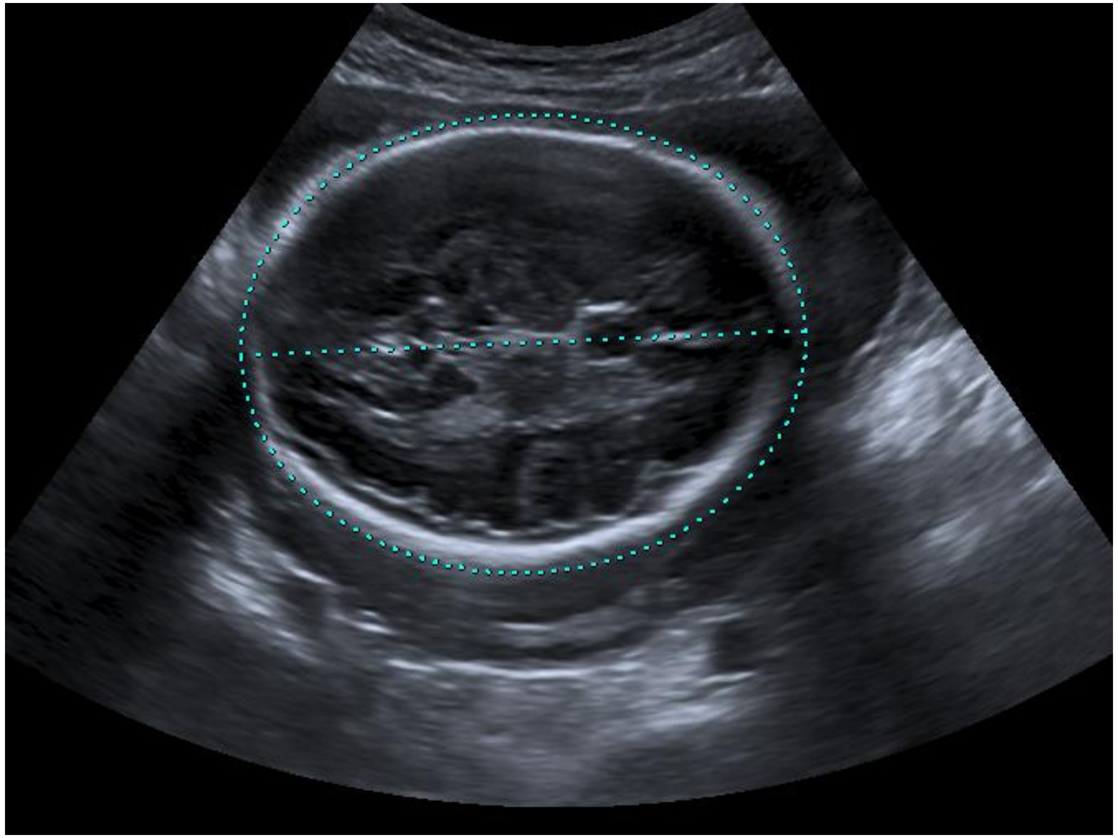
Measurement errors

- Errors due to equipment calibration, resolution etc. very unlikely <u>if you have a QA programme</u>

- Likely sources are:
  - Image plane
  - Measurement method
  - Caliper placement

- May all be influenced by
  - Image quality
  - Training and experience
  - Level of standardisation amongst staff

Example of overmeasurement



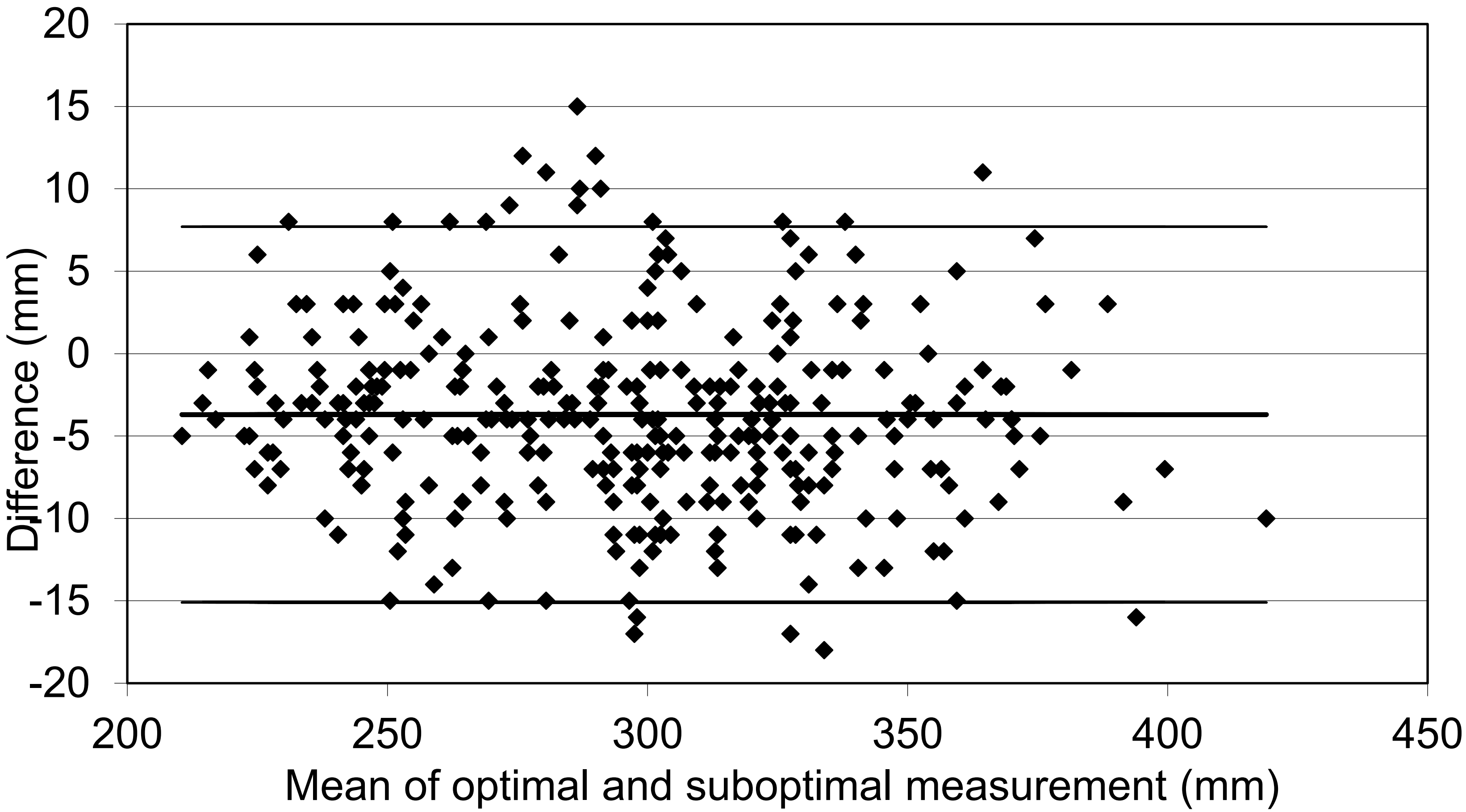
This could represent consistent overmeasurement or inconsistent fitting

Image plane

- Multicentre study auditing against image plane criteria
- HC – 87% met all criteria (78% to 95%)
- AC – 60% met all criteria (45% to 75%)

Dudley NJ, Chapman E. The importance of quality management in fetal measurement. *Ultrasound Obstet Gynecol* 19: 190-196, 2002

Optimal vs sub-optimal plane (AC)



Dudley NJ, Chapman E. The importance of quality management in fetal measurement. *Ultrasound Obstet Gynecol* 19: 190-196, 2002

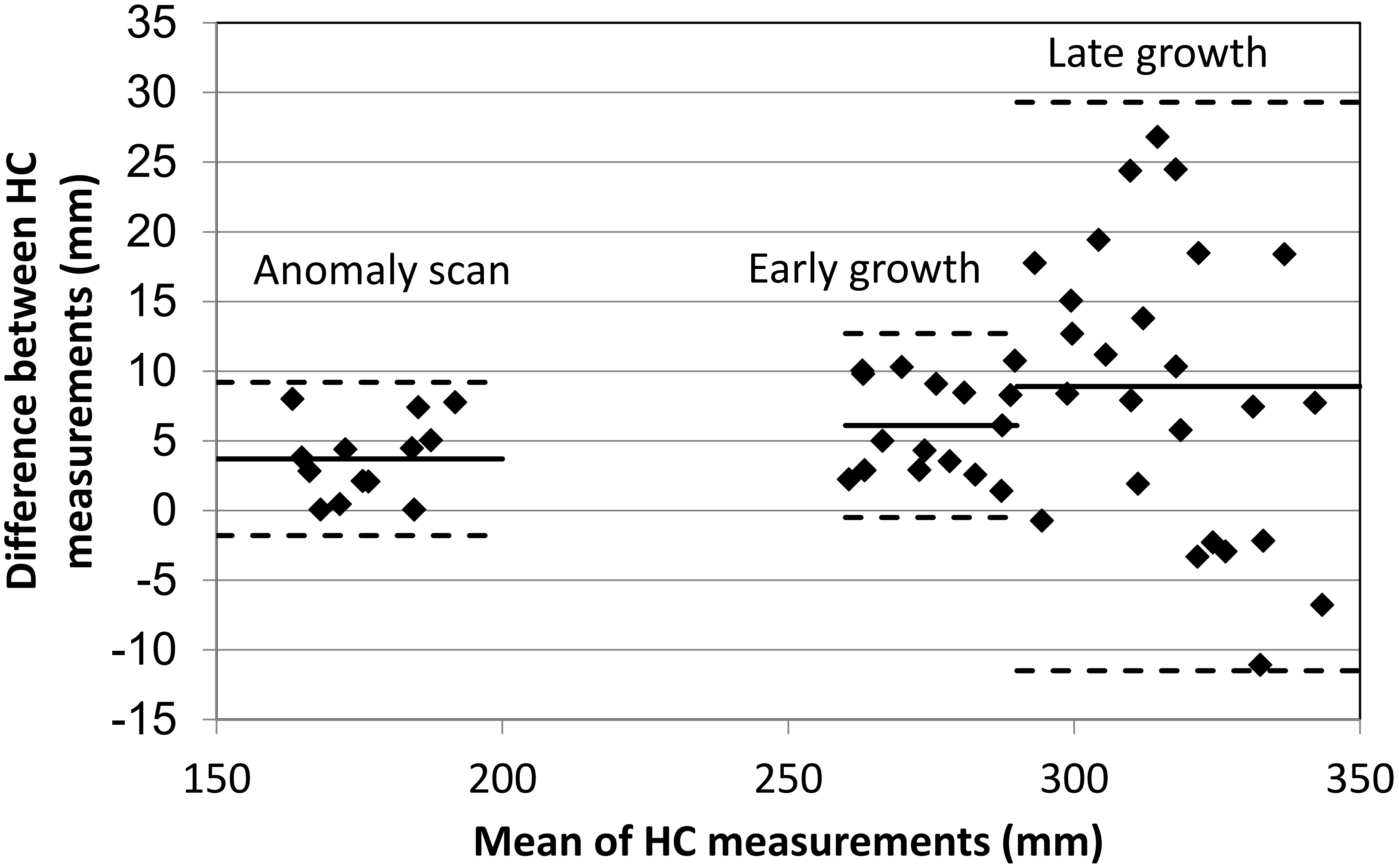
Image plane

Sub-optimal image planes can easily lead to errors of 5% or more in an individual measurement

Measurement method

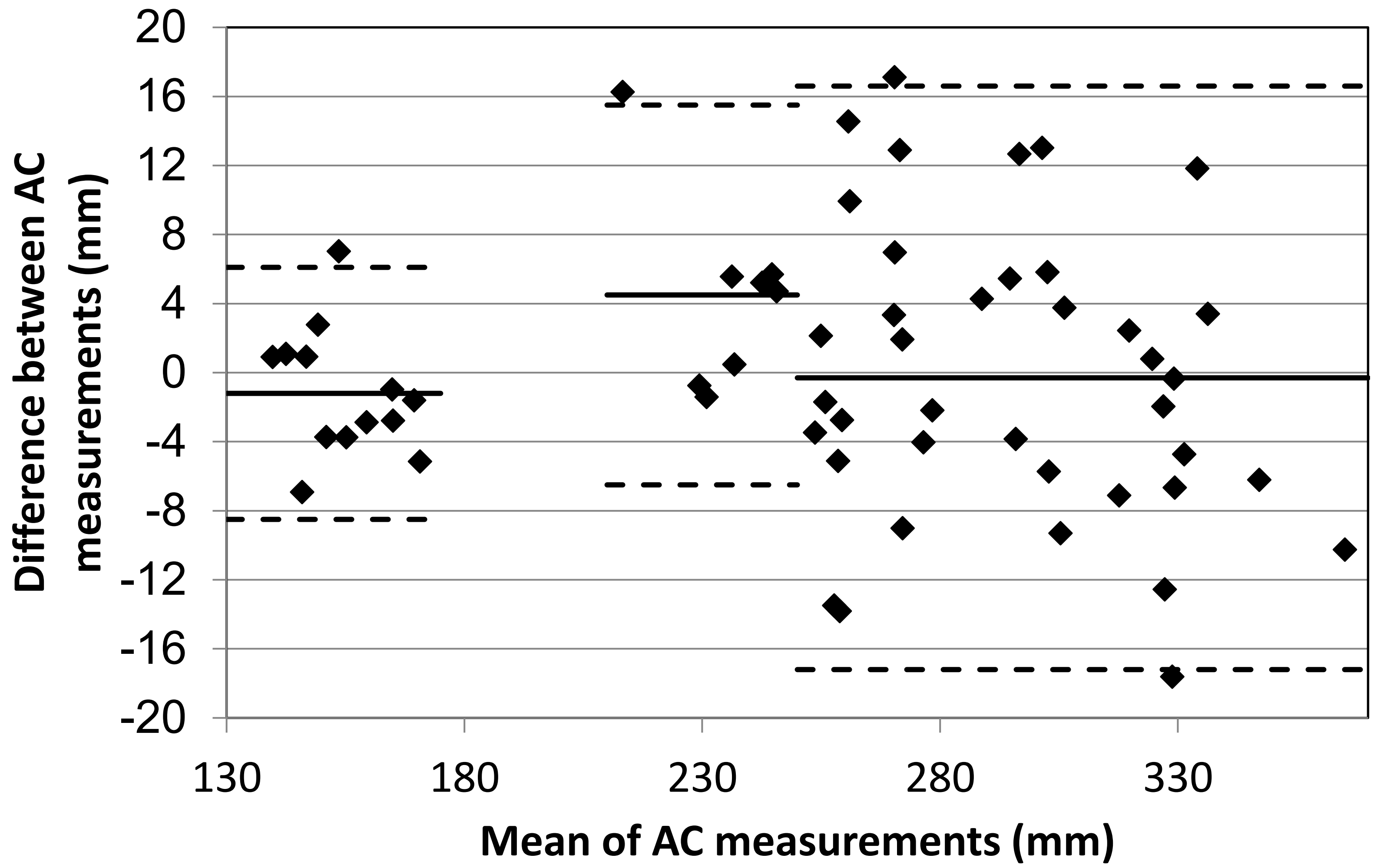
- No choice for FL – linear
- HC and AC – trace, ellipse, cross …
- Cross (2 orthogonal diameters) used/recommended by:
- Hadlock et al.
- Chitty et al.
- BMUS (Chitty charts)
- INTERGROWTH-21st recommend ellipse

Are ellipse/cross interchangeable?



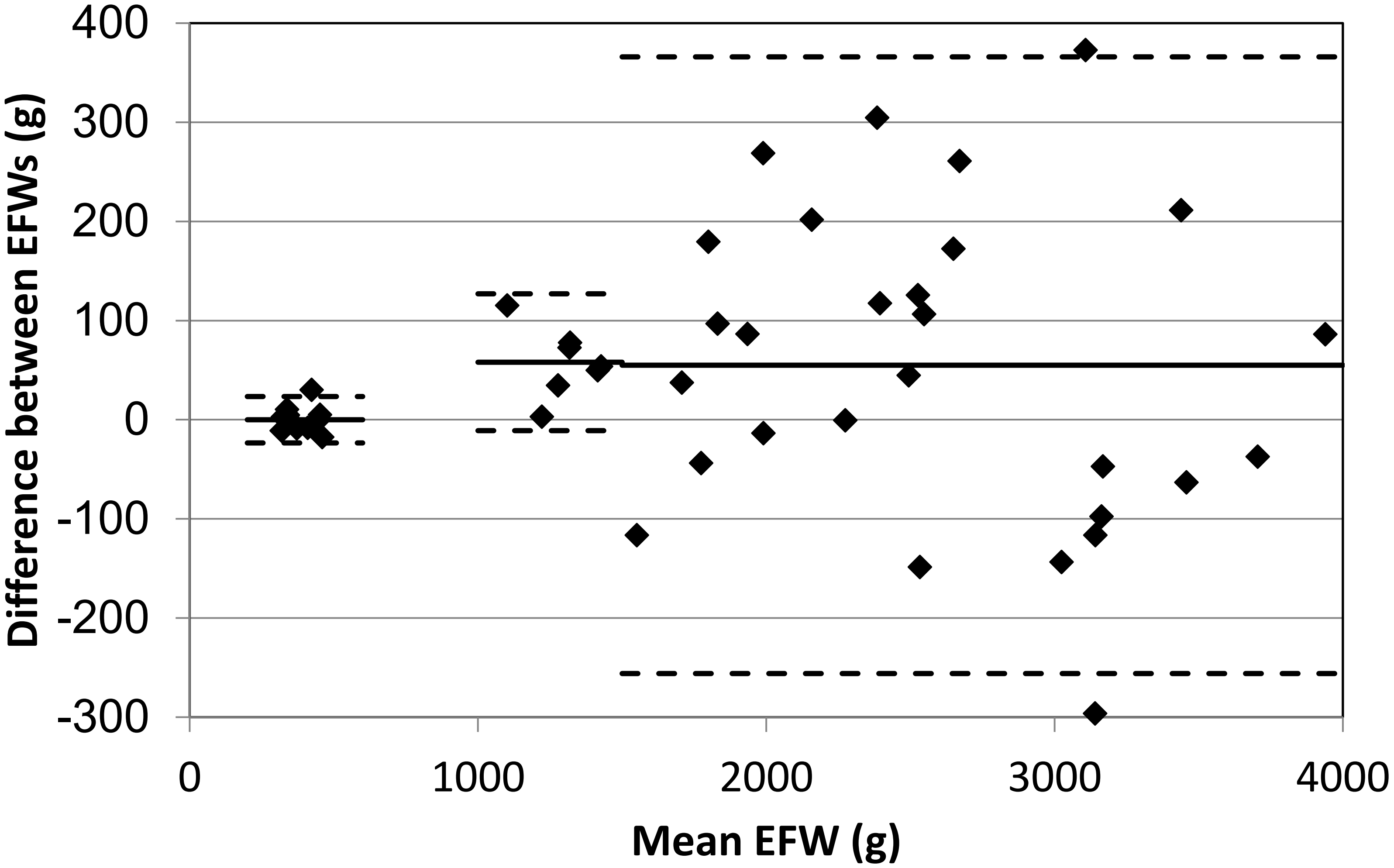
Dudley NJ. Are fetal circumference measurement methods interchangeable? *Ultrasound* 27: 176-182, 2019

Are ellipse/cross interchangeable?



Dudley NJ. Are fetal circumference measurement methods interchangeable? *Ultrasound* 27: 176-182, 2019

Impact on EFW



Dudley NJ. Are fetal circumference measurement methods interchangeable? *Ultrasound* 27: 176-182, 2019

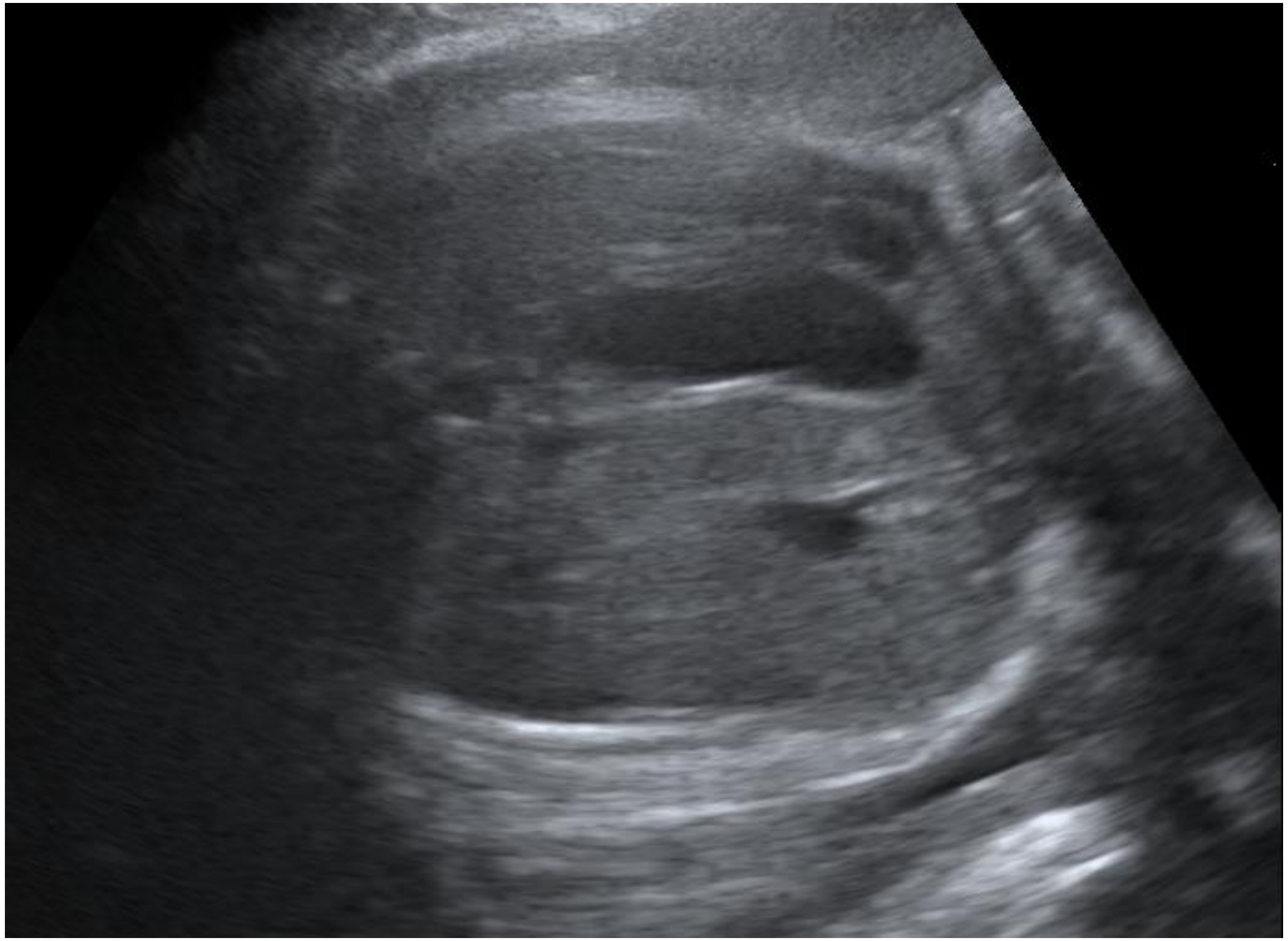
Measurement method

The choice between ellipse and cross can easily generate differences of 10% or more in EFW

Caliper placement

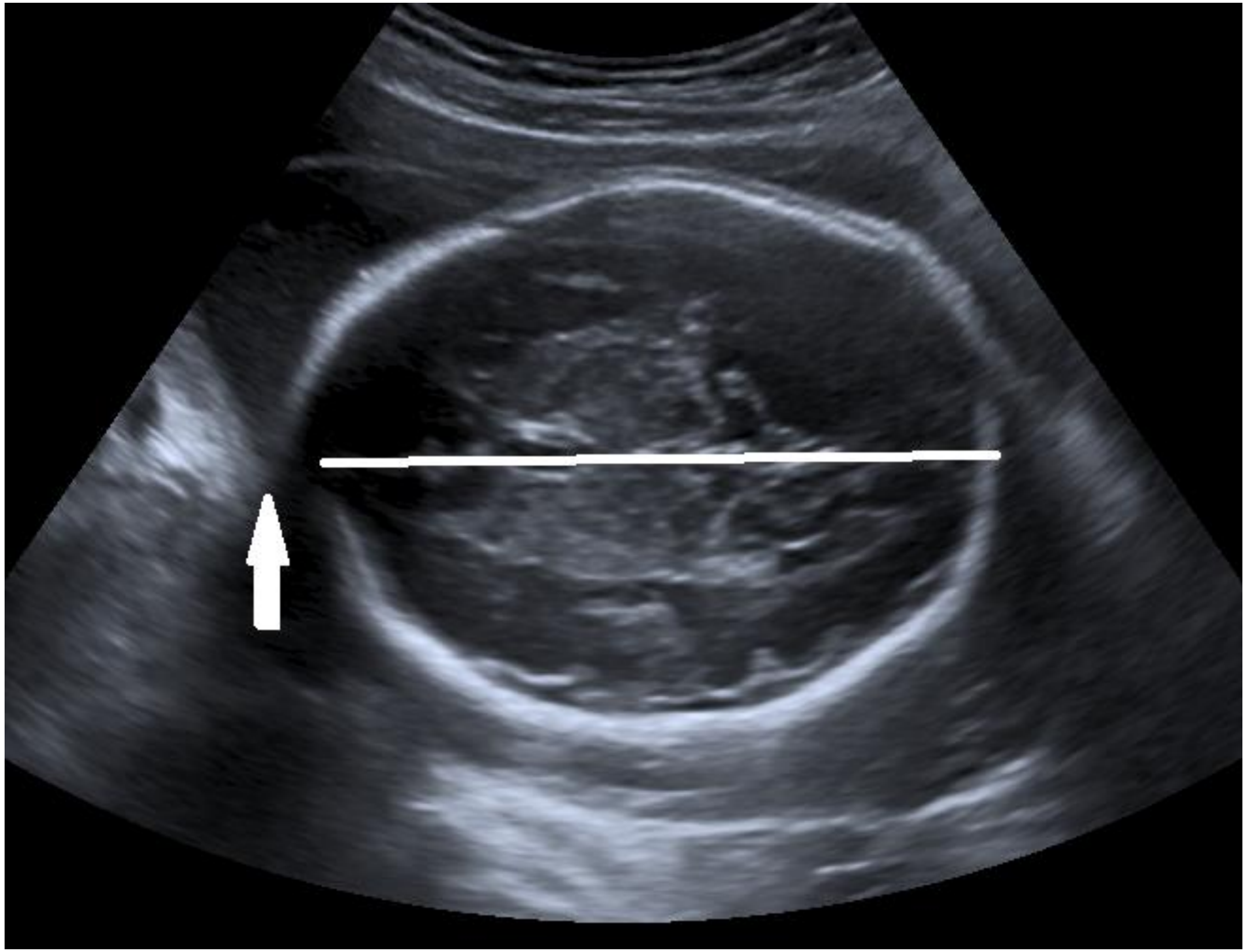
- The critical final stage in making an ultrasound measurement
- Important to match criteria used to develop charts
- Important to carefully place the calipers on the correct end points
  - outer-to-outer
  - skull surface
  - skin surface
  - etc.

Method and calipers



Sometimes it's just difficult!

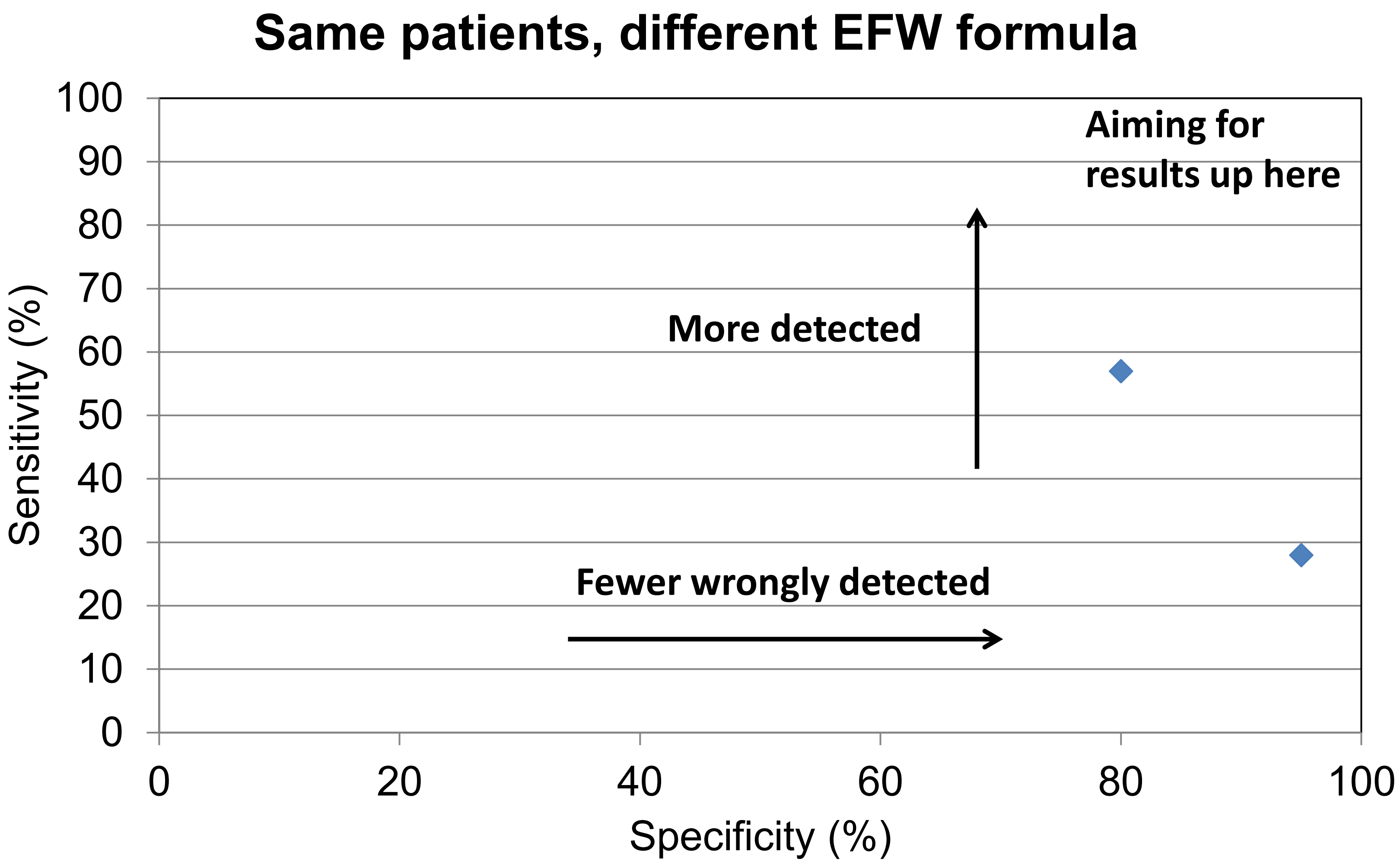
Method and calipers



EFW formula

- There are many!
- They give different results!
- Need to match the measurement methods and the formula to the fetal growth chart.
- Hadlock's AC, HC, FL formula is the method of choice.
- Hadlock's AC, FL formula is used as an alternative, can give a different result.

EFW formulae



Interpretation

- We know errors are potentially large.
- We want to detect every FGR.
- Do we need a high threshold (to include error margin)?
- Threshold at 35th percentile detected100% of SGA births.

Dudley NJ, Lamb MP, Hatfield JA, Copping C & Sidebottom K. Estimated fetal weight in the detection of the small for menstrual age fetus. *J Clin Ultrasound* 18:387-393. 1990

Interpretation - GROW/GAP

- Evidence based method reduces confounding influence of errors.
- Abnormal growth defined to include a margin of uncertainty,
  - e.g. average of -8% versus expected, for slow growth.
- Identifies slow or accelerated growth even for fetuses in AGA range.
- Fetuses in AGA EFW range born with SGA were still relatively small; average centile of 31.5 (cf. 35^{th} percentile threshold in previous slide.

Hugh O, Gardosi J. Fetal weight projection model to define growth velocity and validation against pregnancy outcome in a cohort of serially scanned pregnancies. *Ultrasound Med Biol* 60: 86-95, 2022

Summary of error sources

- Measurements (AC, HC, FL)
  - Literature suggests maybe 50% or more of the EFW error is due to the observer
- EFW formula
  - So … is up to 50% of the error due to the EFW formula?
- Interpretation (charts and procedures)
  - Do we know how to interpret EFW, in the presence of large uncertainty?

Errors in context

Saving Babies Lives Care Bundle v.3 (July 2023)

- 10th and 3rd centiles key thresholds: ~8% difference
- Suboptimal growth definition <420g in 3 weeks <u>removed</u> (~5% of EFW)
- Still a requirement to look at growth velocity and refer to AC growth

EFW errors in the literature (average)

- 1 in 3 errors >12%
- 1 in 20 errors >24%

What we do and don’t know

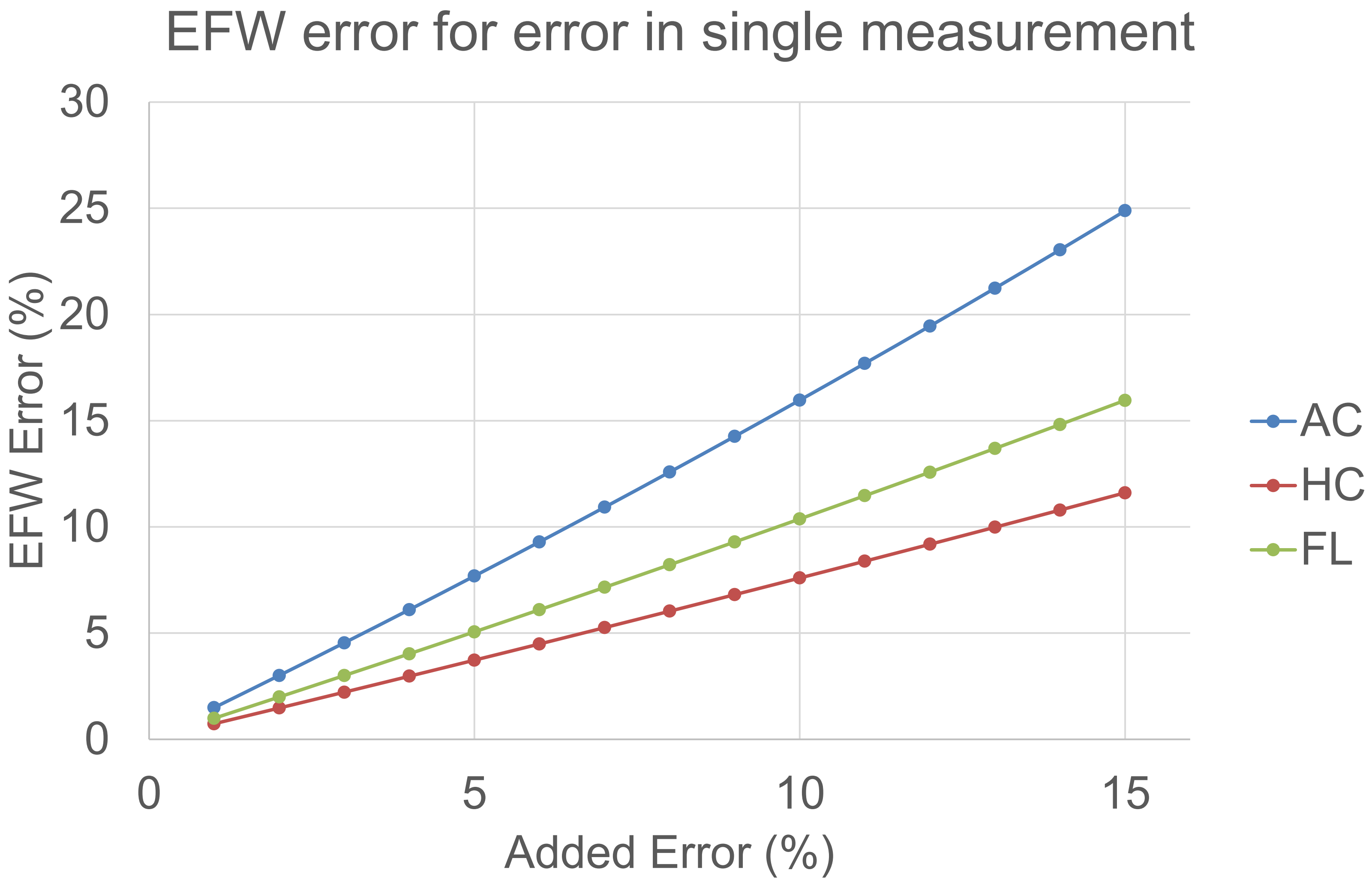
- We know about measurement errors
- We know interpretation is difficult in the presence of uncertainty
- Do we know the contribution of the EFW formula?

Sensitivity of Hadlock (AC, HC, FL) formula to errors

$\log_{10}(EFW) = 1.5662 - 0.0108 \times HC + 0.0468 \times AC + 0.171 \times FL + 0.00034 \times HC^2 - 0.003685 \times AC \times FL$

- How do errors in the individual measurements affect EFW?
- Experiment:
- Used average AC, HC, FL (Chitty charts) at 32 weeks to calculate EFW.
- In turn, added “errors” to AC, HC, FL, 1% to 15%.
- Plotted EFW “error” against errors added to measurements.

Sensitivity of Hadlock (AC, HC, FL) formula to errors



Types of EFW formula

$$\log_{10}(EFW) = 1.5662 - 0.0108 \times HC + 0.0468 \times AC + 0.171 \times FL + 0.00034 \times HC^2 - 0.003685 \times AC \times FL$$

(Hadlock et al.)

$$EFW = 0.32 \times AC^2 \times FL + 0.053 \times HC^2 \times FL$$ (Dudley et al.)

$$EFW = 0.23718 \times AC^2 \times FL + 0.03312 \times HC^3$$ (Combs et al.)

EFW formulae and "model" fetus

- Mathematical model of fetus, verifed with MRI data
- Fetal proportions varied, e.g. shorter/fatter, longer/thinner
- Volume like formulae best able to cope with changes in fetal proportions, within ±5%
  - supported by Mongelli & Biswas (2002)
- Log formulae less robust, errors ±8%
- Formulae contribute to error, via differences in fetal proportions

How can we improve accuracy now?

- Standardisation and procedures
- Training
- Audit

Standardisation and procedures

- Image plane
  - standards widely agreed
- Scanner settings
  - output, frequency, compounding, scale, focus …
- Measurement method
  - ellipse – open to different interpretation, perception of fit
  - cross – less open to interpretation, outer-to-outer
- EFW formula – the right one(s) entered correctly

Training

- Recognition of image plane criteria
- Ability to obtain image
- Correct measurement method

“Input” audit

- Image audit
  - continuous audit too time-consuming
  - “snapshot” audit, smaller number of images
  - feedback and coaching on criteria not met
  - repeat audit
  - maintains high awareness of standards

- Methods
  - several published methods
  - score against image plane criteria, magnification and caliper placement
  - output should be % with all criteria met

HC audit image

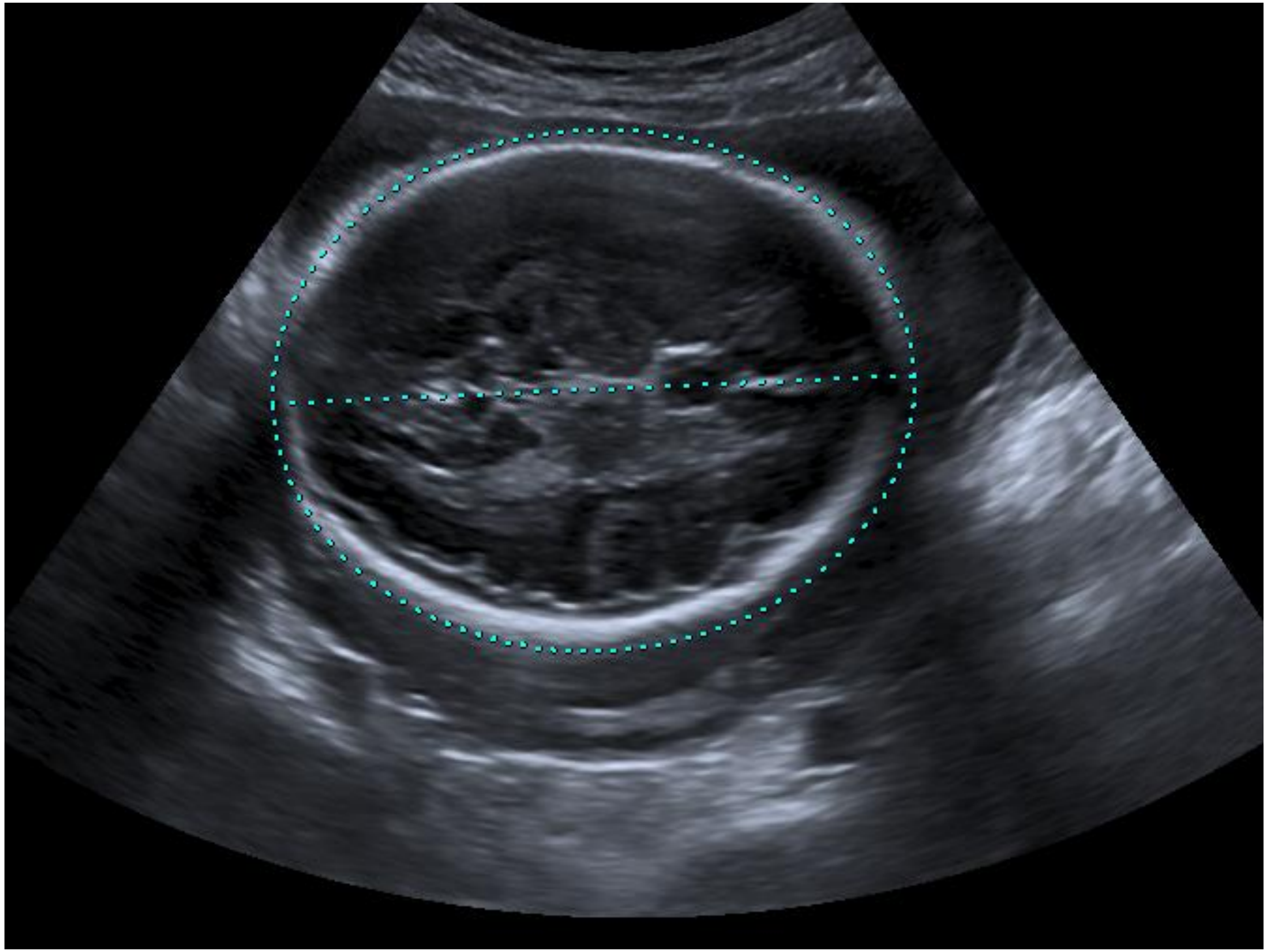
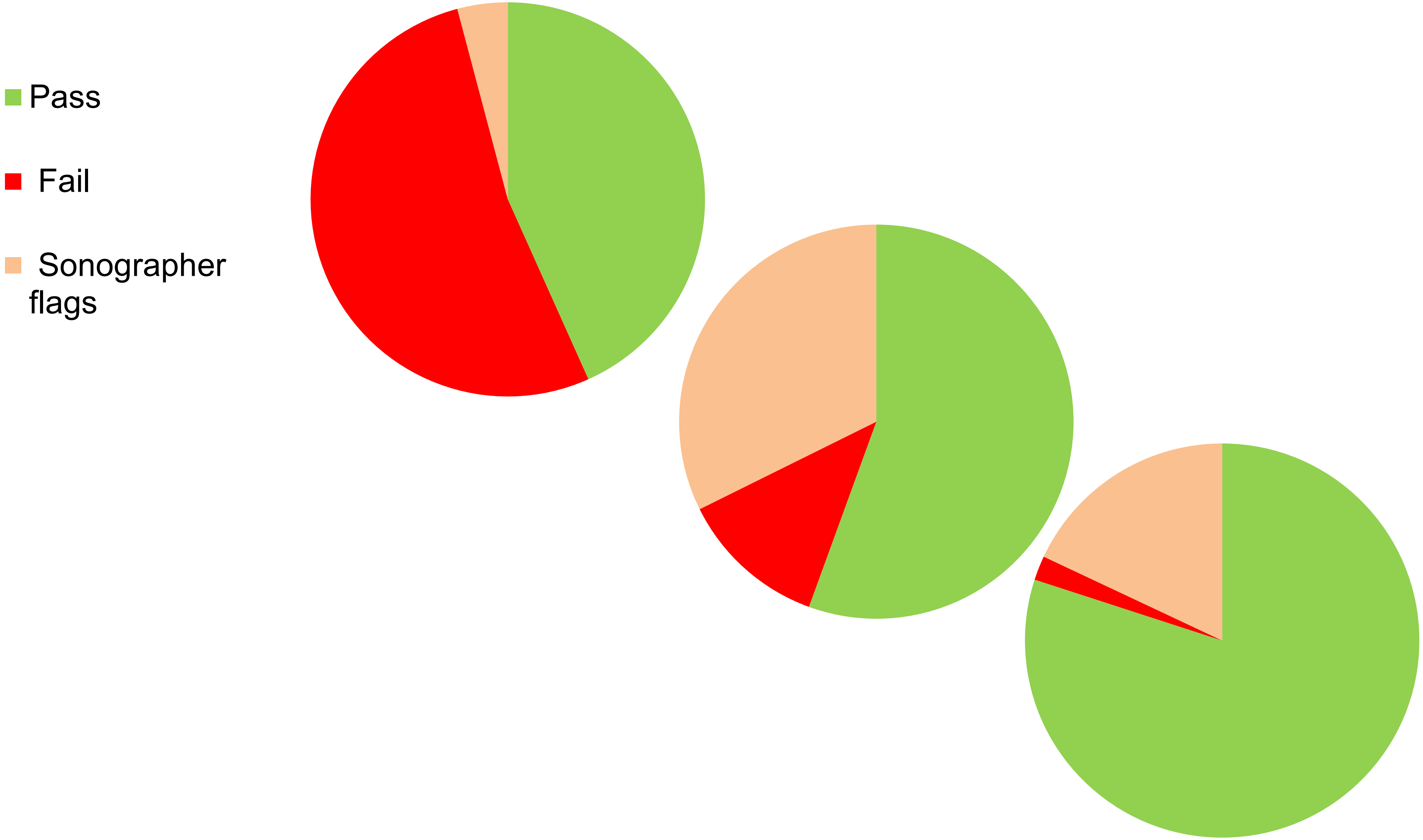


Image audit



Number of images needed?

- BMUS recommendations:

- Audit 5% of all growth scan images.
- Recognise target ambitious so a minimum of 3 randomly selected full measurement sets per person every 3 months.
- 25 of a single measurement audited on a rotational basis.

- What is the uncertainty in the results from 3 or 25 images?

https://www.bmus.org/mediacentre/news/professional-guidance-for-fetal-growth-scans-performed-after-23-weeks-of-gestation/ BMUS 2022

Uncertainty in image audit

- Theoretical outcomes can be calculated using statistics.
- If we toss a coin, the outcome is a head or a tail.
- Expect to see an approximately equal number of heads and tails – the chance of either is 50%, or a probability of 0.5.
- 3 tosses, could get 3 heads, 3 tails, 2 heads and 1 tail, 2 tails and 1 head (4 outcomes).
- Same for pass/fail audit – 4 possible outcomes: 0%, 33%, 67%, 100%.

Dudley NJ, Platten DJ. Ultrasound Fetal Biometry: Number needed to audit. *Ultrasound* Online 2026

Uncertainty in image audit

- If someone throws 3 heads in a row, would we be surprised?
- If someone who gets a good measurement 50% of the time gets 3 good ones (or 3 bad ones) in a row would we be surprised?
- Using statistics, we can calculate the chance of different outcomes from the odds of success/failure.
- Data from historic audits mentioned earlier show a range of pass rates from 45% to 91%; these figures used for the “known” odds of success.

Uncertainty in image audit

Pass result (100 images)	45%	75%	90%
95% CI (3 images)	±56.2%	±49.0%	±33.9%
Potential range (3 images)	0% - 100%	33% - 100%	67% - 100%

Uncertainty in image audit

Pass result (100 images)	45%	75%	90%
95% CI (3 images)	±56.2%	±49.0%	±33.9%
Potential range (3 images)	0% - 100%	33% - 100%	67% - 100%
95% CI (10 images)	±30.8%	±26.8%	±18.6%
Potential range (10 images)	10% - 80%	50% - 100%	70% - 100%

Uncertainty in image audit

Pass result (100 images)	45%	75%	90%
95% CI (3 images)	±56.2%	±49.0%	±33.9%
Potential range (3 images)	0% - 100%	33% - 100%	67% - 100%
95% CI (10 images)	±30.8%	±26.8%	±18.6%
Potential range (10 images)	10% - 80%	50% - 100%	70% - 100%
95% CI (25 images)	±19.5%	±17.0%	±11.8%
Potential range (25 images)	24% - 64%	60% - 92%	80% - 100%

Uncertainty in image audit

Pass result (100 images)	45%	75%	90%
95% CI (3 images)	±56.2%	±49.0%	±33.9%
Potential range (3 images)	0% - 100%	33% - 100%	67% - 100%
95% CI (10 images)	±30.8%	±26.8%	±18.6%
Potential range (10 images)	10% - 80%	50% - 100%	70% - 100%
95% CI (25 images)	±19.5%	±17.0%	±11.8%
Potential range (25 images)	24% - 64%	60% - 92%	80% - 100%
95% CI (100 images)	±9.8%	±8.5%	±5.9%
Potential range (100 images)	35% - 55%	67% - 83%	84% - 96%

“Output” audit

- EFW versus birthweight
  - Perinatal Institute provides a tool
  - growth adjusted EFW from the latest scan vs birth weight
  - assumes “normal” growth between scan and delivery

- How many do we need to be sure that apparent differences in performance are real?

- Sample size calculation:

$$d > 1.96 \times \sigma \times \sqrt{(2/n)}$$

Uncertainty in EFW audit

Number of samples (n)	3	10	25	100
Significant difference	16.8%	9.2%	5.8%	2.9%

- If Sonographer 1 has a mean error of 5% and Sonographer 2 has a mean error of 20% for 3 EFWs each, it could be due to chance.
- The same difference for 10 or more EFWs is statistically significant.
- The more EFWs audited, the smaller the significant difference becomes.

Pragmatic advice for audits (see publication)

- Low number of images:
  - do not focus on, or compare, pass rates for individuals as these have considerable uncertainty;
  - combine results from all individuals and offer feedback to the whole group on the most often failed criteria, without identifying the results of individuals;
  - once you have larger numbers, provide individual feedback.

- Low number of EFWs:
  - do not try to interpret!
  - wait until a large enough number has been collected before analysing and feeding back.

Dudley NJ, Platten DJ. Ultrasound Fetal Biometry: Number needed to audit. *Ultrasound* Online 2026

“Acceptable” Error?

- Known Confidence Intervals are not a measure of acceptability.
- Clearly a 12% error has potential to cause problems with interpretation.
- Acceptability depends on context.
- If investigating a poor outcome, a 15% (or even a 25%) error in EFW may not be due to poor practice.
  - “Forgivable” error, “error within expected margins”, not “acceptable” error.
- In a clinical context 15% is not “acceptable”.

Improving accuracy

- Automated image quality assessment and measurement

- A new EFW formula
  - we are working with formulae that are over 40 years old!
  - (Hadlock et al. 1985)

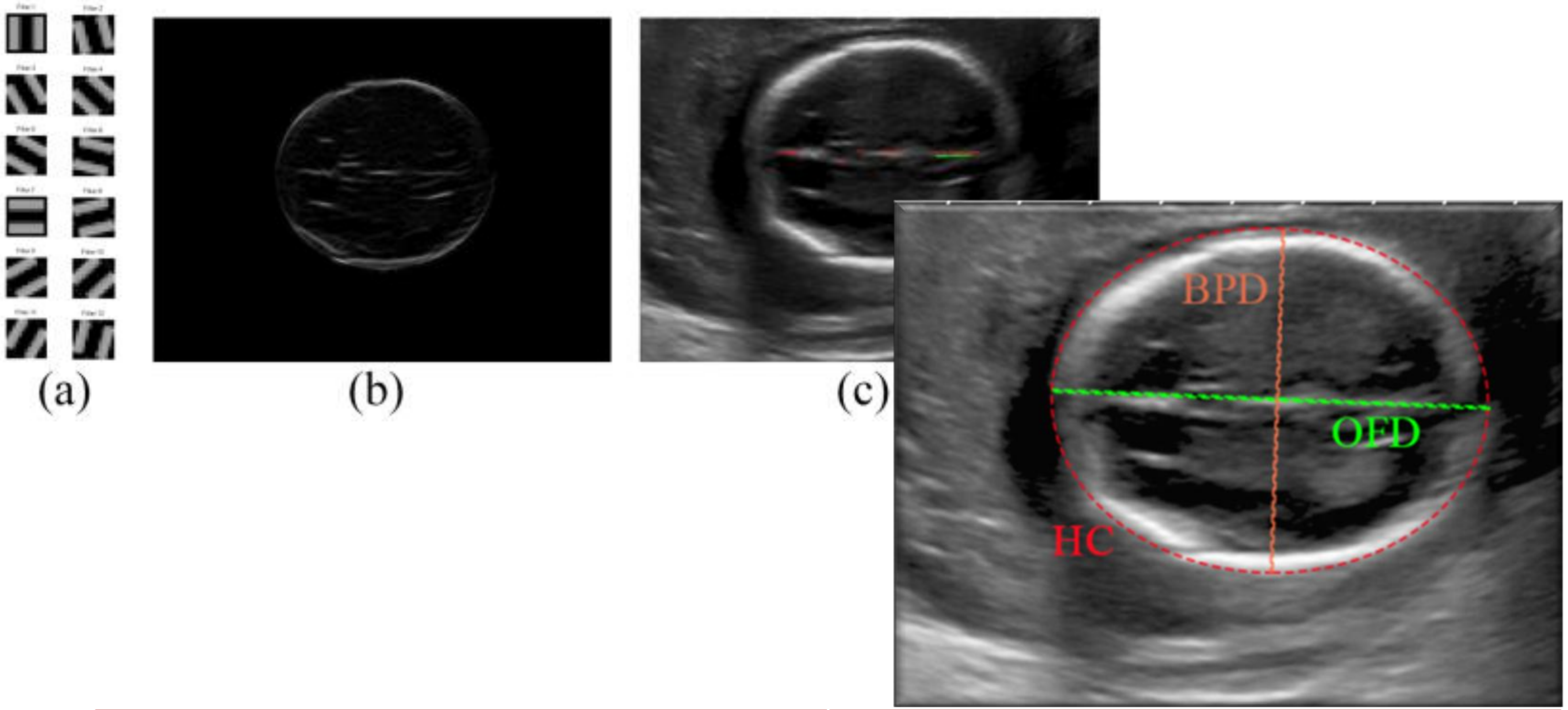
What is AI?

- Machine learning.
- Neural networks.
- Deep learning.
- Not “programmed” as traditional computing.
- Generates associations between input data and required output.

AI training

- Multiple examples of input data, e.g. AC measurement.
- System “learns” how to generate required output.
- Internal processes of AI often unknown – “black box”.
- More advanced systems will recognise/choose optimum plane as well as measuring.
- Identification of plane and measurement are two very different tasks.

Automated image quality assessment and measurement



Parameters	Our method
Accuracy(%)	95.24%
Sensitivity (%)*	87.5%
Specificity (%)*	100%

AI and EFW

- Commercial systems available that recognise and measure images.
- AI has been used to combine measurements and GA to estimate fetal weight.
- Ultimately should reduce variability in measurements,
  - no inter- or intra-observer differences.

Questions to ask before using “AI” measurement

- Is it accurate and reproducible?
- Is the performance better than human, more accurate, faster, both?
- Are measurements interchangeable with manual methods?
- If not, have specific charts and calculations been developed?
- Does it ever need over-riding by the user?
- Are users able to recognise when it’s wrong or sub-optimal?
- …

- Experience with systems that automatically generate a result is that they can be used badly and the result may rarely be questioned.

New EFW formula

- Modelling suggests volume like formula more robust
- Use automated measurements, reduce inherent random errors
- Try to understand errors arising from formula and fetal proportions; customised EFW?
- Appeal for data – I have a new method but no data!

Summary

- EFW errors can be large
  - part formula
  - part observer
  - interpretation important
- Standardisation / Training / Audit
  - image plane, multiple criteria
  - measurement methods not interchangeable
  - EFW formula not interchangeable
  - snapshot audit of images
- The future
  - automated image quality assessment and measurement
  - guided interpretation
  - updated EFW methods

Further reading. Questions welcome: dudleyn5@gmail.com

1. Dudley NJ, Platten DJ. Ultrasound Fetal Biometry: Number needed to audit. *Ultrasound* Online 2026
2. Dudley NJ, Jindal S, Varley H. An evaluation of systematic and random errors in ultrasound estimated fetal weight during serial ultrasound. *Ultrasound* 31: 259-265, 2023
3. Dudley NJ. The management of error in ultrasound fetal growth monitoring. *Ultrasound* 29: 4-9, 2021
4. Dudley NJ, Varley H. Caveats in the monitoring of fetal growth using ultrasound estimated fetal weight. *Ultrasound* 29: 10-17, 2021
5. Dudley NJ. Are ultrasound foetal circumference measurement methods interchangeable? *Ultrasound* 27: 176-182, 2019