Scientific basis for standardization of fetal head measurements by ultrasound: a reproducibility study

R. NAPOLITANO*, V. DONADONO*, E. O. OHUMA*,†‡, C. L. KNIGHT*, S. Z. WANYONYI*, B. KEMP*, T. NORRIS* and A. T. PAPAGEORGHIOU*†

*Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Oxford, UK; †Oxford Maternal & Perinatal Health Institute, Green Templeton College, University of Oxford, Oxford, UK; ‡Centre for Statistics in Medicine, Botnar Research Centre, University of Oxford, Oxford, UK

KEYWORDS: biparietal diameter; fetal head biometry; head circumference; occipitofrontal diameter; reproducibility; transthalamic and transventricular plane; ultrasound; variability

ABSTRACT

Objective To compare the standard methods for ultrasound measurement of fetal head circumference (HC) and biparietal diameter (BPD) (outer-to-outer (BPDoo) vs outer-to-inner (BPDoi) caliper placement), and compare acquisition of these measurements in transthalamic (TT) vs transventricular (TV) planes.

Methods This study utilized ultrasound images acquired from women participating in the Oxford arm of the INTERGROWTH-21st Project. In the first phase of the study, BPDoo and BPDoi were measured on stored images. In the second phase, real-time measurements of BPD, occipitofrontal diameter (OFD) and HC in TT and TV planes were obtained by pairs of sonographers. Reproducibility of measurements made by the same (intraobserver) and by different (interobserver) sonographers, as well as the reproducibility of caliper placement and measurements obtained in different planes, was assessed using Bland–Altman plots.

Results In Phase I, we analyzed ultrasound images of 108 singleton fetuses. The mean intraobserver and interobserver differences were <2% (1.34 mm) and the 95% limits of agreement were <5% (3 mm) for both BPDoo and BPDoi. Neither method for measuring BPD showed consistently better reproducibility. In Phase II, we analyzed ultrasound images of 100 different singleton fetuses. The mean intraobserver and interobserver differences were <1% (2.26 mm) and the 95% limits of agreement were <8% (14.45 mm) for all fetal head measurements obtained in TT and TV planes. Neither plane for measuring fetal head showed consistently better reproducibility. Measurement of HC using the ellipse facility was as reproducible as HC calculated from BPD and OFD. OFD by itself was the least reproducible of all fetal head measurements.

Conclusions Measurements of BPDoi and BPDoo are equally reproducible; however, we believe BPDoo should be used in clinical practice as it allows fetal HC to be measured and compared with neonatal HC. For all head measurements, TV and TT planes provide equally reproducible values at any gestational age, and HC values are similar in both planes. Fetal head measurement in the TT plane is preferable as international standards in this plane are available; however, measurements in the TV plane can be plotted on the same standards. © 2016 The Authors. Ultrasound in Obstetrics & Gynecology published by John Wiley & Sons Ltd on behalf of the International Society of Ultrasound in Obstetrics and Gynecology.

INTRODUCTION

Fetal head biometry is important for estimation of gestational age in the second trimester and for monitoring fetal growth. Unfortunately, even after decades of clinical practice, guidelines still vary as to how the measurements should be taken, i.e. whether the biparietal diameter (BPD) should be measured by outer-to-outer (BPDoo) or outer-to-inner (BPDoi) caliper placement1,2. It is also uncertain whether head circumference (HC) should be calculated from the occipitofrontal diameter (OFD) and BPD (HCcalculated) or by using the ellipse facility (HCellipse) on the ultrasound machine, and which is the better plane to use, i.e. transthalamic (TT) or transventricular (TV)1,3. These issues are important clinically because measurement inconsistencies may affect the management
of individual pregnancies, make it difficult to compare data across units and contribute to the heterogeneity of studies describing fetal size.\textsuperscript{4,5}

In this study, we aimed to compare (i) the standard methods for measuring fetal HC (H_{\text{ellipse}} vs H_{\text{calculated}}) and BPD (BPDoo vs BPDoi caliper placement) on ultrasound and (ii) the effect of acquiring head measurements in TT vs TV planes, so as to make recommendations regarding best practice.

**SUBJECTS AND METHODS**

This study involved women at low risk of adverse pregnancy outcome who were recruited into the Oxford arm of the INTERGROWTH-21\textsuperscript{st} Project (www.intergrowth21.org.uk), a multicenter, multiethnic, population-based project, conducted between 2008 and 2014 across eight countries.\textsuperscript{6} The Fetal Growth Longitudinal Study (FGLS) is one of the three main components of the INTERGROWTH-21\textsuperscript{st} Project, which aimed to construct international standards for fetal growth. All women included in our study were part of the FGLS. In the FGLS, serial two-dimensional ultrasound scans were performed every 5 ± 1 weeks, from 14 + 0 to 41 + 6 weeks' gestation, and images were stored for later analysis. Inclusion criteria for the FGLS were pregnant women with a known, certain last menstrual period, who had regular menstrual cycles and were not taking hormonal contraceptives or breastfeeding in the period, who had regular menstrual cycles and were not pregnant women with a known, certain last menstrual period.

In Phase I, only BPDoo was measured to reduce scanning time. All measurements were obtained in a blinded fashion and were stored on the ultrasound machine and retrieved after completion of the study.

Each sonographer placed the calipers once on each of the four images acquired per participant (i.e. a total of 12 measurements per sonographer for BPDoo, OFD and H_{\text{ellipse}}). Sonographer B repeated the caliper placements on the images acquired by Sonographer A, resulting in a total of 36 measurements. HC was also calculated from BPD and OFD (H_{\text{calculated}}) for each image.

**Measurement and plane definitions**

BPDoo was measured with the intersection of the calipers placed from the outer edge of the proximal calvarial wall to the outer edge of the distal calvarial wall, at the widest part of the skull (Figure 1a). BPDoo was measured with the intersection of the calipers placed from the outer edge of the proximal calvarial wall to the inner edge of the distal calvarial wall (Figure 1b).\textsuperscript{10} OFD was measured with the intersection of the calipers placed from the outer edge of the anterior frontal wall to the outer edge of the distal occipital wall, at the longest part of the skull (Figure 1b). H_{\text{ellipse}} was measured using the ellipse facility, placing the line of the ellipse on the outer border of the skull (Figure 1b).\textsuperscript{2} The TT plane was acquired according to the following conditions: axial view at the level of the thalami with an angle of insonation as close as possible to 90\degree; the head had to be oval in shape, symmetrical, centrally positioned and filling at least 30% of the monitor; the midline echo (representing the falx cerebri) had to be broken anteriorly, at a third of its length, by the cavum septi pellucidi; and the thalami had to be located symmetrically on either side of the midline (Figure 1b).\textsuperscript{2} The TV plane was acquired including all the standard parameters to obtain a TT plane but visualizing the lateral ventricles rather than the thalami at a more cranial level, with the ventricles located symmetrically on each side of the midline, the anterior and posterior horns both visible, and the posterior ventricle cavity visualized as a hypoechoic structure (Figure 1a).\textsuperscript{1}
Figure 1  (a) Ultrasound image of biparietal diameter, measured using outer-to-outer caliper placement (BPDoo), and fetal head circumference (HC), measured using the ellipse facility, in the transventricular plane. (b) Ultrasound image of biparietal diameter, measured using outer-to-inner caliper placement (BPDoi), and occipitofrontal diameter (OFD) in the transthalamic plane.

Statistical analysis

In Phase I, the following analyses were performed: (i) intraobserver reproducibility of caliper placement for measurement of BPD using the BPDoo and BPDoi method, calculated for Sonographers A and B; and (ii) interobserver reproducibility of caliper placement for measurements of BPD using the BPDoo and BPDoi method, comparing the first measurements of Sonographer A with those of Sonographer B, and the second measurements of Sonographer A with those first obtained by Sonographer B.

In Phase II, the following analyses were performed: (i) intraobserver reproducibility of plane acquisition and caliper placement for TT and TV planes, comparing each sonographer’s first and second measurements in the same plane; (ii) interobserver reproducibility of plane acquisition and caliper placement for TT and TV planes, comparing measurements of Sonographers A and B in the same plane; (iii) caliper replacement reproducibility, based on Sonographer B replacing the calipers on the images acquired by Sonographer A in the TT and TV planes (interobserver reproducibility); (iv) intraobserver reproducibility of plane acquisition and caliper placement between TT and TV planes, comparing the measurements of Sonographer A acquired in the TT plane with those acquired by Sonographer B; and (v) interobserver reproducibility for plane acquisition and caliper placement between TT and TV planes, comparing the measurements of Sonographer A acquired in the TT plane with those acquired by Sonographer B in the TT plane and then the measurements of Sonographer B acquired in the TT plane with those acquired by Sonographer A in the TV plane.

Intraobserver and interobserver variability were expressed as a percentage to account for increasing fetal head size with gestational age. Percentages were calculated as the difference between two measurements divided by the average of the two measurements, multiplied by 100.

Reproducibility was assessed using Bland–Altman plots. All plots and analyses were performed using STATA 11 (StataCorp, College Station, TX, USA).

Paired or unpaired t-tests, as appropriate, were performed to assess mean differences between measurements obtained by the same sonographer (intraobserver reproducibility) and different sonographers (interobserver reproducibility), and those obtained in two different planes (between-plane reproducibility). A P-value of < 0.05 was considered statistically significant.

RESULTS

Four women were included in the study at each gestational week, from 15 to 41 weeks in Phase I (108 women) and from 16 to 40 weeks in Phase II (100 women), resulting in a total of 4464 measurements. The demographic characteristics of the 208 participants are shown in Table 1.

Phase I: evaluation of biparietal diameter caliper placement

A total of 864 measurements were obtained in Phase I. Intraobserver and interobserver reproducibility was very good overall. The mean differences were < 2% (1.34 mm) and the 95% limits of agreement were < 5% (3 mm) for both BPDoo and BPDoi (Table 2 and Figures S1 and S2); however, neither method showed consistently better reproducibility. As expected, the 95% limits of agreement for interobserver reproducibility of BPDoo and BPDoi (3.1–4.2%) were slightly wider than for the intraobserver reproducibility (1.3–2.1%).

Phase II: evaluation of transthalamic vs transventricular plane

A total of 3600 measurements (1200 for BPD, OFD and HCellipse) were obtained in Phase II. HCellipse was
Table 1 Demographic characteristics of women with singleton pregnancy recruited into the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project who had retrospective measurement of biparietal diameter (BPD) (Phase I) or real-time measurements of fetal biometry in transthalamic (TT) and transventricular (TV) planes (Phase II).

| Characteristic       | Phase I: BPD study (n = 108) | Phase II: TT/TV study (n = 100) |
|----------------------|-------------------------------|---------------------------------|
| Maternal age (years) | 30 ± 4                        | 30 ± 5                          |
| BMI (kg/m²)          | 23.3 ± 2.7                    | 26.9 ± 3.9                      |
| Nulliparous          | 66 (61)                       | 42 (42)                         |
| GA at scan (weeks)   | 28.1 ± 7.7                    | 28.0 ± 7.2                      |

Data are given as mean ± SD or n (%). BMI, body mass index; GA, gestational age.

Table 2 Intra- and interobserver reproducibility of biparietal diameter measurement using outer-to-outter (BPDoo) or outer-to-inner (BPDoi) caliper placement method

| Measurement          | Sonographer A | Sonographer B | Interobserver |
|----------------------|---------------|---------------|---------------|
| BPDoo                | 0.01 (2.08)   | 0.02 (1.28)   | 1.93 (4.16)   |
| BPDoi                | −0.16 (1.63)  | −0.15 (1.33)  | 0.80 (3.10)   |

Data are given as mean difference (95% limits of agreement (LOA)) in percent. Upper and lower 95% LOA in each case can be calculated as mean difference ± value displayed.

We found no major differences in the reproducibility of caliper placement for measuring BPDoo or BPDoi. Similarly, there was no difference in the reproducibility of measuring HC in the TV or TT planes. Using the ellipse facility (HCellipse) to measure HC was marginally more reproducible than using the two-diameters method (HCcalculated), with the former having interobserver 95% limits of agreement of just below 5% and the latter having interobserver 95% limits of agreement of just above 5%. This is probably due to the contribution of the OFD, which is the least reproducible head measurement in the two-diameters method.

The BPDoi method was used originally because the inner margin of the fetal skull in the distal field was sharper when using static B scanners. However, modern equipment produces a clearer image and so the BPDoi method appears to have no measurable effect on reproducibility (Table 2), even though caliper replacement constitutes up to 60% of the total variability. Therefore, choosing between BPDoo and BPDoi should be for reasons other than trying to reduce error, such as the protocol used (BPDoi) to develop international standards for monitoring fetal growth. Another reason for using BPDoo is that it enables direct comparisons to be made between antenatal and postnatal measurements of HC.

Lastly, neither the TV nor TT plane was found to be consistently associated with better reproducibility. We did find that biometry in the TV plane yielded a very slightly larger HC than that measured in the TT plane. Although this was statistically significant, it was not clinically relevant (< 0.1%, 0.61 mm). Furthermore, when comparing the reproducibility of measuring HC in the TT and TV planes, the difference between sonographers measuring in the same plane was similar to that of the same sonographer measuring in different planes.

© 2016 The Authors. Ultrasound in Obstetrics & Gynecology published by John Wiley & Sons Ltd on behalf of the International Society of Ultrasound in Obstetrics and Gynecology.
Table 3 Intraobserver and interobserver reproducibility of ultrasound measurements of fetal head biometry and caliper replacement in the same plane and between planes

| Measurement      | Within-plane reproducibility | Between-plane reproducibility |
|------------------|-------------------------------|-------------------------------|
|                  | Intraobserver | Interobserver | Caliper replacement | TT interobserver | TV interobserver | Intraobserver | Interobserver |
| BPDoo            | TT            | TV             | TT                | TT            | TV              | TT            | TV              |
| −0.14 (4.05)     | −0.02 (3.43)  | 0.70 (6.65)    | 0.09 (4.78)       | 0.30 (3.16)   | 0.41 (2.69)     | 0.24 (5.63)   | 0.24 (5.84)     |
| OFD              | −0.31 (6.55)  | −0.41 (5.50)   | −0.03 (7.98)      | −0.13 (7.66)  | 0.50 (4.63)     | 0.86 (4.58)   | −0.13 (6.69)    | −0.14 (8.11)   |
| HCellipse        | −0.06 (3.47)  | −0.25 (3.32)   | −0.48 (4.78)      | −0.75 (4.87)  | −0.43 (3.14)    | 0.12 (3.05)   | −0.09 (4.53)    | −0.10 (5.11)   |
| HCalculated      | −0.23 (4.13)  | −0.24 (3.53)   | 0.29 (3.54)       | 0.02 (5.02)   | 0.43 (2.91)     | 0.66 (2.92)   | 0.04 (4.78)     | 0.03 (5.50)    |

Data are given as mean difference (95% limits of agreement (LOA)) in percent. Upper and lower 95% LOA in each case can be calculated as mean difference ± value displayed. BPDoo, biparietal diameter measured using outer-to-outer caliper placement; HCalculated, head circumference calculated from biparietal diameter and occipitofrontal diameter (OFD); HCellipse, head circumference measured using ellipse facility on ultrasound machine; TT, transthalamic; TV, transventricular.

Limitations and strengths

There are some limitations to our study. It can be argued that the use of six different sonographers working in pairs (rather than one pair) might have had an impact on the results. However, we feel that the study design more accurately reflects clinical practice, as most units have several qualified sonographers22. The setting of near-optimal conditions (i.e. experienced sonographers, healthy population and a scientifically rigorous study design) may be seen as creating an artificial setting. However, such conditions were necessary to minimize the contribution of confounding factors so as to define the variability in relation to the research question as purely as possible, which we see as a strength. The other strengths of our study were that reproducibility was assessed throughout pregnancy by recruiting a fixed number of women per week of gestation, and recommended methods23 were used that have been shown to be the most appropriate for assessing the reproducibility of two measurements24,25.

Our findings in context with other studies

A literature search was performed to identify all publications reporting reproducibility in the evaluation of fetal head biometry. We searched MEDLINE using the following keywords: biparietal diameter OR BPD OR occipitofrontal diameter OR OFD or head circumference OR HC AND fetal OR foetal OR fetus OR foetus AND ultrasound OR ultrasonography OR ultra-sonography* OR sonic* OR scan* AND reproducibility OR variability OR repeatability. Restrictions that were applied were studies in humans, in the English language and published after 1970. Additional references were added from an important article4. Nineteen relevant studies were identified (Table S1)15–18,22,26–39. In most, the primary aim of the study was not to assess reproducibility but to build growth charts. The studies reporting either BPD method did not reveal large differences from our findings (the reported mean differences were <2% for BPDoo, with limits of agreement of <5%),15–18,34,36 and there were only two small studies29,38 on BPDoo showing limits of agreement of 3.8 and 7.4 mm, respectively).

In only one study was the reproducibility of BPDoo and BPDoi reported in the same group of fetuses, which showed repeatability coefficients that were similar for both methods34. Measurements of HCellipse were reproducible, with a mean difference of 3.5 mm and limits of agreement of <12 mm (5%), in line with our results15–17,22,27–29,34,35,39. However, BPDoo can be used for both BPD and HC measurements and is also the method to measure OFD. It therefore seems simplest to use BPDoo as a conceptually similar methodological approach for all head measurements. BPDoo is also clinically useful (as part of the HCalculated) for monitoring growth from the ‘womb to the classroom’40, as it is possible to track head size and growth from the antenatal to postnatal periods41. We found that HC measurements using HCellipse were associated with slightly better interobserver reproducibility than using HCalculated, based on BPD and OFD. However, there was no large difference in reproducibility of BPD, OFD or HCellipse measured in the TV compared with TT plane. The mean difference in head size between these two planes was also minimal (<1%) at every gestational age.

We therefore recommend that standard fetal head biometry measurements are performed using the BPDoo, OFD and HCellipse, all measured in the TT plane, based on the reproducibility evidence presented in this study and the existence of international standards based on these methods. In centers in which HC is measured in the TV plane, use of the international standards is still appropriate19.

Acknowledgments

A.T.P. is the Chief Medical Officer of Intelligent Ultrasound and receives non-financial support from Philips Ultrasound. This project was supported by a generous grant from the Bill & Melinda Gates Foundation to the University of Oxford (Oxford, UK), for which we are very grateful.
REFERENCES

1. Sonographic examination of the fetal central nervous system: guidelines for performing the 'basic examination' and the 'fetal neurosonogram'. Ultrasound Obstet Gynecol 2007; 29: 109–116.

2. Papageorghiou AT, Sarris I, Ioannou C, Todros T, Carvalho M, Pilia G, Salomon LJ. Ultrasonography: the fetal growth standards in the INTERGROWTH-21st Project. BJOG 2013; 120 Suppl 2: 27–32.

3. Salomon LJ, Althier Z, Bergella H, Bilardo C, Hernandez-Andrade E, Johnsen SL, Kaulke K, Leung KY, Malinger G, Munoz H, Preftos F, Toi A, Lee W. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. Ultrasound Obstet Gynecol 2011; 37: 116–126.

4. Ioannou C, Talbot K, Othman E, Sarris I, Villar J, Conde-Aguado A, Papageorghiou A. Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size. BJOG 2012; 119: 1425–1439.

5. Ng T, Lhottta R, Dhami J, Othman EA, Ioannou C, Conde-Aguado A, Kennedy SH, Villar J, Papageorghiou AT. Pregnancy dating by fetal crown-rump length: a systematic review of charts. BJOG 2014; 121: 556–565.

6. Villar J, Altman DG, Parwar M, Noble JA, Knight HE, Ruyan P, Cheikh Ismail L, Barros FC, Lambea A, Papageorghiou AT, Carvalho M, Jaffer YA, Bertino E, Gravett MG, Bhutta ZA, Kennedy SH. The objectives, design and implementation of the INTERGROWTH-21st Project. BJOG 2013; 120 Suppl 2: 9–26.

7. Wan L, Wang X, Nolato n, Othman E, Salomon I, Papageorghiou A. Image-scoring system for crown-rump length measurements. Ultrasound Obstet Gynecol 2014; 44: 649–654.

8. Ioannou C, Sarris I, Hoch L, Salomon LJ, Papageorghiou AT. Standardisation of crown-rump length measurement. BJOG 2013; 120 Suppl 2: 38–41.

9. Sarris I, Ioannou C, Othman E, Altman DG, Hoch L, Cosgrove C, Fatima S, Salomon L, Papageorghiou A. Standardisation and quality control of ultrasound measurements taken in the INTERGROWTH-21st Project. BJOG 2013; 120 Suppl 2: 33–37.

10. Leung TN, Pang MW, Daljit SS, Leung TY, Poon CF, Wong SM, Lau TK. Fetal biometry in ethnic Chinese: biparietal diameter, head circumference, abdominal circumference and femur length. Ultrasound Obstet Gynecol 2008; 31: 321–327.

11. Ioannou C, Javid MK, Mahon P, Yaquf MK, Harvey NC, Goodfry KM, Noble JA, Cooper C, Papageorghiou AT. The effect of maternal vitamin D concentration on the development of the fetal brain. J Clin Ultrasound 2012; 39: E2070–E2077.

12. Lo Zito L, Kadij C, Janiny M, Kacem Y, Strizk R, Mbonyumutwa M, Wuyts F, Toi A, Lee W. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. Ultrasound Obstet Gynecol 2011; 37: 116–126.

13. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med 1990; 20: 371–376.

14. Ioannou C, Chamberlain P, Othman E, Roseman F, Hoch L, Altman DG, Papageorghiou AT. Intra- and interobserver variability in fetal ultrasound measurements. Ultrasound Obstet Gynecol 2012; 39: 266–273.

15. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1: 307–310.

16. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ 1992; 304: 1491–1494.

17. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Ultrasound Obstet Gynecol 2000; 15: 3–10.

18. Natesan S, Natesan R, Karras A, Bhutta ZA, Kennedy SH, Barros FC, Carvalho M, Altman DG, Salomon LJ, Papageorghiou A. Longitudinal distance standards of fetal growth. Intrauterine and Infant Growth Standards: the INTERGROWTH-21st Project. 2013; 48: 556–565.

19. Hoffman FP, Deter RL, Harrist RB, Park SK. Fetal head circumference: relation to fetal age, sex, and gestational age. Obstet Gynecol 1987; 78: 347–351.

20. Hadlock FP, Deter RL, Harrist RB, Park SK. Fetal head circumference: relation to menstrual age. AJR Am J Roentgenol 1982; 138: 649–653.

21. Johns a SL, Wilgaard T, Rasmussen S, Soll en R, Kiserud T. Longitudinal reference charts for growth of the fetal head, abdomen and femur. Eur J Obstet Gynecol Reprod Biol 2006; 127: 172–185.

22. Krampl E, Lees C, Bland JM, Everett DC, Mosevo s G, Campbell S. Fetal biometry at 4300m compared to sea level in Peru. Ultrasound Obstet Gynecol 2000; 15: 9–18.

23. Larsen T, Petersen S, Greisen G, Larsen JF. Normal fetal growth evaluated by longitudinal ultrasound examinations. Early Hum Dev 1999; 54: 37–43.

24. Luna JC, Miragaya AH, Filho FP, Zanini CO, Martin VRP. Biometry and fetal weight estimation by two-dimensional and three-dimensional ultrasonography: an intraobserver and interobserver reliability and agreement study. Ultrasound Obstet Gynecol 2012; 40: 116–126.

25. Merialdi M, Caufield LE, Zavaleta A, Figueroa F, Campbell S. Fetal head circumference by gestational age and sex: the Newborn Cross-Sectional Study (INTERGROWTH-21st). International standards for newborn weight, length, and head circumference by gestational age. The likeness of fetal growth and newborn size across the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). The likeness of fetal growth and newborn size across non-isolated populations in the INTERGROWTH-21st Project: The Fetal Growth Longitudinal Study and Newborn Cross-Sectional Study. Lancet Diabetes Endocrinol 2014; 2: 781–792.

SUPPORTING INFORMATION ON THE INTERNET

Table S1 and Figures S1–S7 may be found in the online version of this article.