Meta-analysis of sex difference in testosterone levels in umbilical cord blood.

JA Barry[1], PJ Hardiman[2], MR Siddiqui[3], M Thomas[4].

[1] Department of Psychology, City University, London EC1V OHB.

[2] Institute of Women's Health, University College Medical School, London NW3 2PF

[3] Department of Colorectal Surgery, St Mark's Institute, London, United Kingdom.

[4] Department of Clinical Biochemistry, Royal Free Hospital, London, United Kingdom.

Correspondence:

John A. Barry

Department of Psychology

City University,

London EC1V 0HB.

Email j.a.barry@city.ac.uk

Telephone 020 7040 5060; Fax 020 7040 8580

Keywords: testosterone, umbilical, assay, sex difference, meta-analysis, polycystic ovary syndrome.

Running title: Sex difference in umbilical cord testosterone

Summary

This meta-analysis reviewed published literature comparing human male and female umbilical cord total testosterone (T) levels. Eighteen studies using 1229 samples from 602 male and 627 female newborns were analyzed using the *RevMan 5* statistical package. Analysis using the inverse variance method based on a random-effects model revealed significantly higher cord T in boys than girls at a moderate effect size (Hedges' $g = 0.57$). There was significant heterogeneity between the 18 studies, though the five studies using direct assays showed no heterogeneity. For studies using extraction and chromatography, those that combined T from arterial and venous cord blood found a larger sex difference than those using only cord venous samples (Hedges' $g = 0.94$ *versus* 0.32); this suggests umbilical cord venous T is of maternal/placental origin and arterial T is of fetal origin. The wide range of T values between studies suggests high cross-reactivity in the assay methods reviewed.

Introduction

Barker (2004) proposed that some diseases of adulthood may have their origin in conditions in the fetal environment, and that low birth weight is associated with insulin resistance, type 2 diabetes, hypertension, and coronary heart diseases in adults. This hypothesis is supported by research showing that elevated maternal testosterone (T) is related to low birth weight in sheep (Manikkam et al., 2006) and humans (Carlsen et al., 2006). Elevated maternal T in rhesus monkeys in late pregnancy causes hyperinsulinemia, and in early pregnancy additionally causes type II diabetes (Abbott et al., 2009). Elevated maternal T in either early or late gestation doubles the risk of the female offspring of rhesus monkeys suffering from anovulation and polycystic ovaries in adulthood (Resko et al., 1987). Animal research suggests that females may be more sensitive to the effects of T on brain development than males (Roselli et al., 2007) so it is especially important to assess potential risks to the female fetus of exposure to T. However at present there is no definitive consensus as to whether T is normally lower at birth in the umbilical cord of healthy human

females pregnancies than in male, so it has been difficult to make clinical judgements regarding prenatal risk of T exposure based on levels of umbilical cord T. It is hoped that the identification of normative sex differences in T levels in the present meta-analysis will contribute to our scientific and clinical understanding of this issue.

Polycystic ovary syndrome (PCOS) affects 5-10% of women (Franks, 1995) and is hypothesised to develop as a result of exposure to elevated T prenatally (Dumesic et al., 2006). Some recent research suggests that umbilical vein T is elevated in female newborns of mothers with PCOS (Barry et al., 2010). Placental aromatase is traditionally thought to protect the fetus from raised maternal T, but animal research contradicts this hypothesis (Resko et al., 1987; Manikkam et al., 2006; Abbott et al., 2009).

To date, studies have produced mixed evidence regarding a sex difference in umbilical vein T. Some studies have found that compared to female newborns, male newborns have more umbilical vein T (Herruzo et al., 1993), less (Pardo et al., 1993), or virtually identical levels (van de Beek et al., 2004). Establishing a norm for umbilical cord T is difficult because the commonly used 'direct' assay methods (described in discussion section, below) lack the sensitivity to accurately detect T at levels below 10 nmol/l (Rosner et al., 2007); although umbilical vein T tends to be higher than circulating T

4

in women of reproductive age (Baik et al., 2004) it is usually below the 10 nmol/l detection threshold suggested for direct assays (see Table II).

Fetal T can be measured from various sites. Although amniotic fluid is the best candidate to investigate the effects of early fetal androgen exposure (van de Beek et al., 2004), it is not comparable to cord T at the end of gestation because amniotic T is sampled around weeks 11 to 21 when a sex difference in T is likely to largest because of the testosterone surge in male fetuses (Smail et al., 1981). Most studies of fetal T sample from the umbilical vein at birth; fewer studies sample mixed (arteries + vein) cord blood, and very few sample directly from umbilical arteries. Almost all studies measure total T rather than free (unbound) T.

Some papers on this topic (Mathur et al., 1980; Bolton et al., 1989; Pardo et al., 1993; Troisi et al., 2003; Anderson et al., 2010) speculate or infer that umbilical arterial T is of fetal origin and umbilical vein T is of maternal/placental origin. This hypothesis has not been proved, but if male fetuses produce more T than females, and if the umbilical arteries carry hormones of fetal origin, then the umbilical arteries of males should contain more T than the arteries of females (i.e. a sex difference in arterial T). By extension, because mixed cord T contains arterial T, mixed cord T should show a larger sex difference than the sex difference for venous T.

5

The objective of the present paper was to discover whether there is a sex difference in umbilical vein T across comparable studies..

Materials and Methods

Literature search

All studies that measured testosterone in umbilical cord listed in Pubmed and Medline published up to March 1st 2010, and EMBASE from 1980 to March 31st 2010, were identified. The Cochrane Reviews database was also searched. The keyword search terms, 'umbilical' 'cord' and 'testosterone', were entered simultaneously. This produced 115 articles from Pubmed from 1965 to 2010, and 73 from Medline. Medline produced no new papers in addition to those cited in Pubmed. EMBASE produced four additional studies, but these did not meet the inclusion criteria for the present meta-analysis. The "related article" function was used to widen the results. Additionally three Mesh searches were performed using firstly the terms "testosterone" AND "umbilical cord", "testosterone" AND "umbilical veins", and "testosterone" AND "umbilical

arteries". This retrieved 52, 26, and 16 publications respectively, all of which were previously found using the Pubmed keyword search. The Cochrane Reviews database produced six publications, but these were not relevant. A hand search of relevant articles referenced in these publications were obtained, which produced nine publications not previously found. Each article was assessed by [*authors 1 and 4*], and articles that fitted the main criteria (assaying T in umbilical cord vein in healthy pregnancies) were accessed. Methodological quality was assessed based on the criteria of the Scottish Intercollegiate Guidelines Network (2010).

Inclusion and exclusion criteria

Human studies that compared male and female umbilical cord total T at birth were eligible for inclusion provided that:

a. The pregnancies were healthy

b. The deliveries were spontaneous, or else planned caesarean sections

c. The assay methods were IA, CLIA/ECL, a method using extraction, or mass spectrometry.

Papers with titles or abstracts that indicated that they were not relevant (for various reasons e.g. reviews, single case studies etc) were excluded. A literature search flow chart (S1), an additional Forest plot (S2), table of excluded studies (S3), conversion table for T to nmol from other units (S4), and tables

of methodological quality (S5a, b, and c) are available as supplemental digital content on the *JOG* website **[insert URL].**

Table 1 shows the characteristics of the included studies.

[Table I. here]

Of the included studies, six used direct methods (five IA and one CLIA/ECL), eleven used extraction methods (9 with thin layer chromatography and two with di-ethyl ether extraction alone), and one used LCMS. There were no mixed cord X-IA studies, and only one direct assay mixed cord study. Two studies (Dawood & Saxena, 1977; Bolton et al., 1989) measured T in umbilical arteries and vein separately. The venous samples for these two studies were included in the venous subgroup, and the means of their arterial and venous samples combined were included in the mixed cord subgroup. Only the venous samples for these two studies were included in the 'all groups' analysis. Two studies used 'predominantly' (Abramovich, 1974) or 'mainly' (Maccoby et al., 1979) venous blood, so were classified as venous. One study did not state whether they differentiated between arteries and vein (Gol et al., 2004) and was classified as 'mixed'. Thirteen of the studies were classified as being of moderate quality, four were classified as poor (Forest et al., 1973; Abramovich, 1974; Dawood & Saxena, 1977; Shinkawa et al., 1983) and one classified as good (Troisi et al., 2003).

Statistical analysis

Statistical analyses were performed using Review Manager, Version 5 (RevMan 5). Heterogeneity tests suggested that a random effects model was appropriate to assess the sex difference in umbilical vein T levels in seven of the eight subgroups and the studies as a whole. The I^2 value in the direct assay subgroup was zero thus suggesting no problem with heterogeneity, but in the interests of not risking an underestimation of the heterogeneity between studies the more conservative random effects model was used rather than fixed effects. The inverse variance method was used. The effect size was measured as the standard mean difference, calculated using Hedges' *g*. By convention, like Cohen's *d* the thresholds for small, moderate and large Hedges' *g* effect sizes are 0.2, 0.5, and 0.8 respectively. All T values are presented in nmol/l, and were converted from other units for most studies.

Results

Seventy studies of umbilical cord T were retrieved from the electronic databases. Eighteen studies using 1229 samples (602 male and 627 female) qualified for review according to the inclusion criteria. 834 samples were venous (410 male and 424 female), 395 were mixed (192 male and 203 female), and 41 were arterial (Bolton et al., 1989), (Dawood & Saxena, 1977) (21 male and 20 female). Fifty-two trials were excluded.

Table II shows cord T levels (nmol/l) for the included studies. It can be seen that most studies (16 of 18) found higher cord T in boys than girls, and that overall this difference was of a moderate effect size (Hedges' $g = .57$).

[Table II. here]

Table III shows the results of meta-analyses. Although there was a lot of heterogeneity in the findings (evidenced by the large I^2 values) the various groups based on assay types and

sources of serum all indicate significantly higher T in male umbilical cord than female.

[**Table III. here**]

Figure 1 shows the Forest plot of sex difference in umbilical vein T for all included studies. The findings of the studies tend towards the right hand side of the vertical zero point, indicating the tendency of the studies to find higher T in the umbilical cord of boys.

[Figure 1 here]

Figure 2 compares the magnitude of sex differences in venous compared to mixed cord blood in studies using chromatography with extraction. The findings in the lower Forest plot (mixed cord blood) are further to the right of the zero point than those in the upper Forest plot (venous cord blood); this indicates that T values for boys in the lower plot showed a greater sex difference than T values for boys in the higher plot.

[Figure 2 here]

Discussion

Analysis of all eighteen studies combined, and all subgroups, revealed significantly higher umbilical cord T in boys than girls. There was significant heterogeneity between the 18 studies and in all subgroups except for the direct assay venous subgroup. The confidence intervals for the subgroups were generally narrow and showed effect sizes of clinical interest. With the exception of the venous chromatographic extraction subgroup the confidence intervals did not encompass a zero value, which suggests that overall the sex differences were representative of those likely to be seen in the general population of newborns. Four of the 18 studies were rated as poor, largely due to these papers lacking information regarding methodological quality rather than explicitly being of poor quality. The four 'poor' ratings were on the borderline for scores for moderate quality and excluding them from the

overall analysis made almost no difference to the results, thus their inclusion is appropriate.

The most widely available assay methods, in descending order of accuracy, use mass spectrometry, thin layer chromatography (TLC), extraction with ethyl ether, and direct measurement (Rosner et al., 2007). A therefore unexpected finding of this meta-analysis is that, judging by heterogeneity measures, the direct method may be a more reliable indicator than extraction methods of the sex difference in cord T. However it is likely that the direct method assay detected more substances than T alone, and the interference of other steroids and substances may have simply masked the heterogeneity evidenced in the extraction group. The cross-reacting substances are most likely to have been androgens such as 11-keto-testosterone, 11-beta-OH-testosterone, and dihydrotestosterone (Roche Diagnostics, 2000); for this reason the direct assay might be viewed as an omnibus measure of androgens rather than simply a measure of T. This interpretation remains to be confirmed by a superior assay method, though it is of interest that Legro et al (2010) recently found good correlations between T levels measured using a direct assay and using liquid chromatography mass spectrometry in women with PCOS. Only one study to date (Anderson et al., 2010) has measured T levels in umbilical cord samples using liquid chromatography mass spectrometry.

Anderson et al found that 14 healthy girls had nonsignificantly higher mixed cord T than seven healthy boys (0.66±1.01 vs 0.49±0.35 nmol/l). Sixty-six percent of these samples (10 of the 14 female and four of the seven male) had T levels lower than the detection limit of 0.24 nmol/l. These samples were assigned a value of 0.24 nmol/l. The fact that the actual T level is unknown for 66% of these samples indicates that more sensitive mass spectrometry assays are needed.

Two of the 18 studies (11% of the studies) found that girls had higher cord T than boys, and across studies the large amount of variation in the observed T levels gives cause for concern. Rosner et al. (2007) found that some of their direct assays of T in healthy women were roughly 10 times higher than others; the present authors found that two studies (Gol et al., 2004), (Furuhashi, 1982) reported T values over 100 times higher than some other studies using comparable assays and sample sites. These studies were of moderate methodological quality, and the disparity is not explained by other features of the studies. Although the Hedges *g* values of the two studies are in keeping with the other sixteen studies, their relatively high observed T values are suggestive of the unreliability of these assay types. With the exception of one study (Furuhashi, 1982), the extraction methods showed generally lower mean T values with a smaller range than the direct assay studies. Because the studies included in this meta-analysis are similar in most

relevant characteristics, it might be concluded that although the direct assays give reliable findings in terms of the size of the sex difference, no method appears to yield reliable findings in terms of the absolute T levels, especially the direct immunoassay due to its poor specificity, cross reactivity with other steroids and matrix effects (Rosner et al., 2007).

A finding of potential clinical importance is that, using comparable assays, a larger sex difference was seen in mixed cord samples ($g = .94$, a large effect size) than venous samples ($g = .32$, a small effect size). This in turn would suggest that the higher level in males is due at least part to fetal production. This has implications for understanding the etiology of conditions such as PCOS in which prenatal T exposure is theorised to be a causal factor. Future studies might compare T levels in the umbilical arteries and veins in newborns of women with PCOS, and compare these to healthy pregnancies; relatively high T in the umbilical arteries compared to the vein in PCOS, and higher umbilical arterial T in PCOS compared to healthy pregnancies, would suggest fetal T production in this condition. A recent study of PCOS and metformin - a medication that lowers T – had the potential to address this issue, but interpretation of the findings is difficult because the results were not presented for girls whose mothers had PCOS and were not taking metformin (Carlsen & Vanky, 2010)

The conclusion of this meta-analysis is that although there appears to be a reliable sex difference in umbilical vein T of a moderate effect size, the direct and extraction assay methods lack ecological validity; differences in steroid milieu outside that usually used for in vitro diagnostic purposes produce wide differences in results where such assays are employed to measure specific steroids in real-world samples. Current direct assays are known to be poorly discriminative of testosterone at low values (Rosner et al., 2007). This being the case, our knowledge of gross T levels in cord blood remains limited, and norms for umbilical cord T at birth remain to be established through further research using more specific methods, such as tandem mass spectrometry. Nevertheless the findings of the present meta-analysis suggest that serum in the male umbilical cord at birth might contain a higher level of androgen than seen in the female cord at birth, and this might provide clinicians a rough index (i.e. relative to values seen in the opposite sex) as to whether a female newborn has experienced elevated androgen prenatally, or whether a male newborn has experienced reduced androgen prenatally. Future research might also compare sex hormones in both umbilical veins and arteries (not mixed) from male and female progeny as way of identifying the source (maternal/placental, or fetal) of hormones.

Legend on supplemental digital content

S1. Literature search graph

S2. Forest plot of sex differences in umbilical cord venous T

for direct assay methods only (n=5).

S3. Table of excluded studies.

S4. Table for conversion of testosterone to nmol/l from other units.

S5a-c. Tables of methodological qualities of the prospective studies included (Adapted from the Scottish Intercollegiate Guidelines Network, 2010).

References

Abbott DH, Tarantal AF, Dumesic DA. 2009. Fetal, infant, adolescent and adult phenotypes of polycystic ovary syndrome in prenatally androgenized female rhesus monkeys. American Journal of Primatology 71: 776-84.

Abramovich DR. Human sexual differentiation--in utero influences. 1974. Journal of Gbstetrics and Gynaecology of the British Commonwealth 81: 448-53.

Anderson H, Fogel N, Grebe SK, Singh RJ, Taylor RL, Dunaif A. 2010. Infants of women with polycystic ovary syndrome have lower cord blood androstendione and estradiol levels. Journal of Clinical Endocrinology & Metabolism 95: 2180-2186

Baik I, Liu Q, Sturgeon S, Stanek EJ 3rd, Okulicz W, Hsieh CC. 2006. Reproducibility of assays for steroid hormones, prolactin and insulin-like growth factor-1 in umbilical cord blood. Paediatric and Perinatal Epidemiology 20: 79-86.

Barker DJ. The developmental origins of chronic adult disease. 2004. Acta Pædiatrica Supplement 93:26-33.

Barry JA, Kay AR, Navaratnarajah R, Iqbal S, David AL, Bamfo JEAK, Hines M, Hardiman PJ. 2010. Umbilical vein testosterone in female infants born to mothers with Polycystic Ovary Syndrome is elevated to male levels. Journal of Obstetrics & Gynaecology 30: 444-6.

Bolton NJ, Tapanainen J, Koivisto M, Vihko R. 1989. Circulating sex hormone-binding globulin and testosterone in newborns and infants. Clinical Endocrinology 31: 201-7.

Carlsen SM, Jacobsen G, Romundstad P. 2006. Maternal testosterone

    levels during pregnancy are associated with offspring size at

    birth. European Journal of Endocrinology 155: 365-70.

Carlsen SM, Vanky E. 2010. Metformin influence on hormone levels

    at birth, in PCOS mothers and their newborns. Human

    Reproduction 25: 786-90.

Dawood MY, Saxena BB. 1977. Testosterone and dihydrotestosterone

    in maternal and cord blood and in amniotic fluid. American

    Journal of Obstetrics and Gynecology 129: 37-42.

Dumesic DA, Abbott DH, Padmanabhan V. 2006. Polycystic ovary

    syndrome and its developmental origins. Reviews in Endocrine

    & Metabolic Disorders 8: 127-41.

Forest MG Cathiard AM, Bertrand JA. 1973. Evidence of testicular

    activity in early infancy. Journal of Clinical Endocrinology &

    Metabolism 37: 148-51.

Forest MG, Sizonenko PC, Cathiard AM, Bertrand J. 1974.

    Hypophyso-gonadal function in humans during the first year of

    life. 1. Evidence for testicular activity in early infancy. Journal

    of Clinical Investigation 53: 819-28.

Franks S. Polycystic ovary syndrome. 1995. New England Journal of

    Medicine 333: 853-861.

Furuhashi N, Fukaya T, Kono H, Tachibana Y, Shinkawa O, Takahashi T. 1982. Correlation of birth weights with umbilical cord serum LH-hCG, FSH, beta-hCG, Estradiol, Cortisol and Testosterone levels. Gynecologic and Obstetric Investigation 13: 241-8.

Gol M, Altunyurt S, Cimrin D, Guclu S, Bagci M, Demir N. 2004. Different maternal serum hCG levels in pregnant women with female and male fetuses: does fetal hypophyseal--adrenal--gonadal axis play a role? Journal of Perinatal Medicine 32: 342-5.

Herruzo AJ, Mozas J, Alarcón JL, López JM, Molina R, Molto L, Martos J. 1993. Sex differences in serum hormone levels in umbilical vein blood. International Journal of Gynecology & Obstetrics 41: 37-41.

Legro RS, Schlaff WD, Diamond MP, Coutifaris C, Casson PR, Brzyski RG, Christman GM, Trussell JC, Krawetz SA, Snyder PJ, Ohl D, Carson SA, Steinkampf MP, Carr BR, McGovern PG, Cataldo NA, Gosman GG, Nestler JE, Myers ER, Santoro N, Eisenberg E, Zhang M, Zhang H; Reproductive Medicine Network. 2010. Total testosterone assays in women with polycystic ovary syndrome: precision and correlation with hirsutism. Journal of Clinical Endocrinology & Metabolism 95: 5305-13.

Maccoby EE, Doering CH, Jacklin CN, Kraemer H. 1979.

    Concentrations of sex hormones in umbilical-cord blood: their

    relation to sex and birth order of infants. Child Development

    50: 632-42.

Maffeis C, Moghetti P, Vettor R, Lombardi AM, Vecchini S, Tatò L.

    1999. Leptin concentration in newborns' cord blood:

    relationship to gender and growth-regulating hormones.

    International Journal of Obesity and Related Metabolic

    Disorders 23: 943-7.

Manikkam M, Steckler TL, Welch KB, Inskeep EK, Padmanabhan V.

    2006. Fetal programming: prenatal testosterone treatment leads

    to follicular persistence/luteal defects; partial restoration of

    ovarian function by cyclic progesterone treatment.

    Endocrinology 147: 1997-2007.

Mathur RS, Landgrebe S, Moody LO, Powell S, Williamson HO.

    1980. Plasma steroid concentrations in maternal and umbilical

    circulation after spontaneous onset of labor. Journal of Clinical

    Endocrinology & Metabolism 51: 1235-8.

Miyamoto U. 1981. A sex difference of the concentrations of

    gonadotropins, its subunits and sex steroids in cord veins. Acta

    Obstetrica et Gynaecologica Japonica 33: 711-3.

Pardo IM, Geloneze B, Tambascia MA, Pereira JL, Barros Filho AA. 2004. Leptin as a marker of sexual dimorphism in newborn infants. Jornal de Pediatria 80: 305-8.

Penny R, Parlow AF, Frasier SD. 1979. Testosterone and estradiol concentrations in paired maternal and cord sera and their correlation with the concentration of chorionic gonadotropin. Pediatrics 64: 604-8.

Resko JA, Buhl AE, Phoenix CH. 1987. Treatment of pregnant rhesus macaques with testosterone propionate: observations on its fate in the fetus. Biology of Reproduction 37: 1185–91.

Roche Diagnostics. 2000. Elecsys Testosterone Product Information. Available at http://www.roche-diagnostics.ch/resource.php?id=Resourcefile-37084379e9754cf58 Retrieved 5th May 2010.

Roselli CE, Stadelman H, Reeve R, Bishop CV, Stormshak F. 2007. The ovine sexually dimorphic nucleus of the medial preoptic area is organized prenatally by testosterone. Endocrinology 148: 4450-7.

Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. 2007. Position statement: Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. Journal of Clinical Endocrinology & Metabolism 92: 405-13.

Scottish Intercollegiate Guidelines Network (SIGN) guidelines. 2010.

Available from:

http://www.sign.ac.uk/guidelines/fulltext/50/checklist3.html

Retrieved 4th June 2010.

Smail PJ, Reyes FI, Winter JSD Faiman C. 1981. The fetal hormone
environment and. its effect on the morphogenesis of the genital.
System. In Kogan SJ, Hafez ESE (Eds.), Pediatric Andrology.
The Hague: Martinus Nijhoff; pp. 9–19.

Shinkawa O, Furuhashi N, Fukaya T, Kono H, Tachibana Y,
Takahashi T, Suzuki M. 1983. A study on testosterone
secretion in neonates. Nippon Sanka Fujinka Gakkai Zasshi 35:
266-8.

Simmer HH, Frankland MV, Greipel M. 1972. Neutral C_{19}-steroids
and steroid sulphates in human pregnancy. VI. Quantification
of plasma testosterone in cord venous blood. Steroids 19: 215-
28

Simmons D, France JT, Keelan JA, Song L, Knox BS. 1994. Sex
differences in umbilical cord serum levels of inhibin,
testosterone, oestradiol, dehydroepiandrosterone sulphate, and
sex hormone-binding globulin in human term neonates.
Biology of the Neonate 65: 287-94.

Troisi R, Potischman N, Roberts JM, Harger G, Markovic N, Cole B,

    Lykins D, Siiteri P, Hoover RN. 2003. Correlation of serum

    hormone concentrations in maternal and umbilical cord

    samples. Cancer Epidemiology, Biomarkers & Prevention 12:

    452-6.

van de Beek C, Thijssen JH, Cohen-Kettenis PT, van Goozen SH,

    Buitelaar JK. 2004. Relationships between sex hormones

    assessed in amniotic fluid, and maternal and umbilical cord

    serum: what is the best source of information to investigate the

    effects of fetal hormonal exposure? Hormones & Behavior 46:

    663-9.

Table 1 Study characteristics.

Study	Assay	T units	Mother	Delivery	Newborn	T source
Simmer et al (1972)[27]	C-X-IA	"ng %"	Not stated	Vaginal, term	Normal,	Vein
Forest et al (1973)[25]	C-X-IA	ng/100ml	Not stated	Normal, vaginal	Normal	Mixed
Forest et al (1974)[28]	C-X-IA	ng/100ml	Healthy	Normal, vaginal	Normal	Mixed
Abramovich et al (1974)[22]	IA	ng/100ml	Not stated	Mix of C-S and vaginal	Not stated	Mainly venous
Dawood & Saxena (1977)[21]	X-IA	pg/ml	Some amnio.	Spontaneous vaginal	Not stated	Vein
Maccoby et al (1979)[23]	C-X-IA	ng/ml	Not stated	Term, no C-S	Normal	Mostly venous
Penny et al (1979)[29]	C-X-IA	ng/100 ml	Healthy	Vaginal	Not stated	Mixed
Miyamoto (1981)[30]	IA	ng/ml	Not stated	normal pregnancy & delivery	Term	Vein
Furuhashi et al (1982)[31]	C-X-IA	ng/ml	Not stated	Normal, vaginal	Normal	Vein
Shinkawa et al (1983)[26]	IA	ng/dl	Not stated	Not stated	Term	Vein
Bolton et al (1989)[19]	C-X-IA	nmol/l	Not stated	Not stated	Term	Vein
Herruzo et al (1993)[10]	IA	ng/ml	Not stated	Uneventful	Term	Vein
Simmons et al (1994)[32]	X-IA	nmol/l	Healthy	Normal	Not stated	Vein
Maffeis et al (1999)[33]	CLIA	nmol/l	Uncomplicated	Uncomplicated	Term	Vein
Troisi et al (2003)[17]	C-X-IA	ng/dl	Healthy	Normal: SVD or C-S	Not stated	Mixed
Van de Beek et al (2004)[12]	C-X-IA	nmol/l	Healthy	SVD	Term	Vein
Gol et al (2004)[24]	IA	ng/ml	Healthy	Uncomplicated, C-S only	Term	Mixed (?)
Anderson et al (2010)[15]	LCMS	ng/dl	Healthy	74% SVD; 26% C-S; uncomplicated.	>35weeks	Mixed

IA = Direct immunoassay

CLIA = Direct chemiluminescence immunoassay

X-IA = IA, after extraction

C-X-IA = IA, after extraction and chromatography (TLC and/or column)

LCMS = Liquid Chromatography-Mass Spectrometry.

C-S = Caesarian section

Table 2 Testosterone levels (nmol/l) for all of the included studies, in chronological order. '*g*' indicates Hedges' *g*, the standard mean difference between male and female umbilical cord T. Levels for 41arterial samples[19, 21] are not shown.

	Male			Female			Difference
	Mean	**SD**	**n**	**Mean**	**SD**	**n**	**Hedges' *g* [95% CI]**
Simmer et al (1972)[27]	0.69	0.45	16	0.80	0.52	24	-0.22 [-0.85, 0.42]
Forest et al (1973)[25]	1.17	0 .33	35	0.92	0.26	46	0.85 [0.39, 1.31]
Forest et al (1974)[28]	1.24	0.36	51	0.93	0.28	53	0.96 [0.55, 1.36]
Abramovich et al (1974)[22]	2.92	0.59	20	2.63	1.25	20	0.29 [-0.33, 0.91]
Dawood & Saxena (1977)[21]	0.79	0.47	11	0.31	0.15	18	1.51 [0.65, 2.36]
Maccoby et al (1979)[23]	0.97	0.24	58	0.74	0.15	58	1.14 [0.75, 1.54]
Penny et al (1979)[29]	1.35	0.30	21	0.90	0.25	22	1.60 [0.91, 2.30]
Miyamoto (1981)[30]	16.69	3.78	55	14.44	3.23	57	0.64 [0.26, 1.02]
Furuhashi et al (1982)[31]	682.93	43.65	37	659.99	162.2	35	0.19 [-0.27, 0.66]
Shinkawa et al (1983)[26]	7.67	4.23	45	6.80	2.01	44	0.26 [-0.16, 0.68]
Bolton et al (1989)[19]	0.34	0.10	12	0.28	0.37	12	0.21 [-0.59, 1.02]
Herruzo et al (1993)[10]	22.17	11.45	27	14.23	5.62	25	0.86 [0.29, 1.43]
Simmons et al (1994)[32]	2.10	0.3	62	1.80	0.60	63	0.63 [0.27, 0.99]
Maffeis et al (1999)[33]	10.40	5.54	48	8.50	4.24	50	0.38 [-0.02, 0.78]
Troisi et al (2003)[17]	1.01	0.83	49	0.76	0.59	37	0.34 [-0.09, 0.77]
Van de Beek et al (2004)[12]	4.01	2.50	19	3.72	3.30	18	0.10 [-0.55, 0.74]
Gol et al (2004)[24]	875.13	121.2	29	809.2	100.2	31	0.59 [0.07, 1.11]

Anderson et al (2010)[15]	0.49	0.35	7	0.66	1.01	14		-0.19 [-1.10, 0.72]
Subtotal (95% CI)			**602**			**627**		**0.57 [0.37, 0.77]**

Table 3 Results of meta-analysis for all studies and the subgroups

Group	Number of studies	Hedges' g [95% CI]	Z (p)	Chi² (p)	I²
All	18	0.57 [0.37, 0.77]	5.66 (P < 0.00001)	45.96 (P = 0.0002)	63%
All venous	12	0.50 [0.26, 0.73]	4.17 (P < 0.0001)	28.48 (P = 0.003)	61%
All mixed [*]	8	0.81 [0.48, 1.15]	4.77 (P < 0.00001)	17.38 (P = 0.02)	60%
All extraction	11	0.66 [0.36, 0.95]	4.31 (P < 0.0001)	36.79 (P < 0.0001)	73%
Venous extraction	7	0.51 [0.09, 0.92]	2.40 (P = 0.02)	24.06 (P = 0.0005)	75%
All C-X-IA	9	0.59 [0.24, 0.94]	3.31 (P = 0.0009)	33.03 (P < 0.0001)	76%
Venous C-X-IA	5	0.32 [-0.23, 0.86]	1.15 (P = 0.25)	18.76 (P = 0.0009)	79%
Mixed C-X-IA [†]	5	0.94 [0.53, 1.34]	4.51 (P < 0.00001)	11.31 (P = 0.02)	65%
Venous direct [±]	5	0.48 [0.27, 0.68]	4.62 (P < 0.00001)	3.98 (P = 0.41)	0%

95% CI = 95% confidence intervals

C-X-IA = IA after extraction and chromatography (TLC and/or column)

[*] Includes the mean of umbilical arterial T and umbilical vein T combined, for two studies.[19, 21]

[†] Includes the mean of umbilical arterial T and umbilical vein T combined for one study.[19]

[±] Using the fixed variance method improved results slightly (Z = 4.79, P < 0.00001; g = 0.52 [0.31, 0.73]; Chi² = 3.78, (P = 0.44); I² = 0%).

Table 4. Exclusion criteria and excluded studies (N=115) presented in order of frequency and chronologically.

Studies of T (or aromatase etc) function (also usually not presented by sex) (N=24)	Cairrão et al (2008)
	Jin et al (2007)
	Perusquía et al (2007)
	Jin et al (2005)
	Zapata et al (2005)
	Yildiz et al (2005)
	Ijiri et al (2003)
	Zhang et al (2002)
	Nie et al (2001)
	Du et al (2001)
	Cid et al (1994)
	Loganath et al (1992)
	Gunasegaram (1991) (EMBASE)
	Higano et al (1989)
	Milewich et al (1987)
	Lewis et al (1986)
	Sybulski et al (1975)
	Ahluwalia et al (1974) (cocaine users)
	Swartz et al (1974)
	Stárka et al (1974)
	Simmer et al (1972)
	Rosenfield (1971)
	Heyns & De Moor (1971)
	Kobayashi et al (1969)
Animal studies (n=17)	Huang et al (2010)

Hayashi et al (1997)

Okamoto et al (1996)

Heyns et al (1993)

Mitchell et al (1986)

Vreeburg et al (1986)

Ford et al (1980)

Ellinwood et al (1980)

Resko (1977)

Tseng et al (1975)

Mongkonpunya et al (1975)

Challis et al (1974)

Resko (1974)

Resko (1973)

Milgrom et al (1973)

Goy & Phoenix (1972)

Dang and Meusy-Dessolle (1970) (EMBASE)

One sex only, or T not presented separately for each sex (N=17)	Hickey et al (2010)
	Whitehouse et al (2010)
	Whitehouse et al (2010)
	Hickey et al (2009)
	Rohrmann et al (2009)
	Troisi et al (2008)
	Nagata et al (2007) (also arteries only)
	Nagata et al (2006) (also arteries only)
	Baik et al (2006)
	Baik et al (2005)
	Schubring et al (1998)

	Sakai et al (1991) (same sex twins)
	Milewich et al (1990)
	Bradshaw et al (1986)
	Tapanainen et al (1984) (EMBASE)
	Tapanainen (1983)
	Mathur et al (1980)
Total T not measured or not reported (N=12)	Toth et al (2009)
	Clifton et al (2007)
	Tan & Tan (2001) (unclear if total or free T)
	Fausett et al (1999)
	Maffei et al (1998)
	Gemer et al (1997) (maternal T only)
	Adeyemo & Jeyakumar (1993) (Free T only)
	Ikegawa (1986) (Free T only)
	Tojo (1981)
	Plotti et al (1975)
	Nunez et al (1974)
	Ermini et al (1974) (T sulphate)
Same data from another study (N=12)	Faupel-Badger et al (2009) [Troisi et al's 2006a data]
	Savage (2009) (Cohrane cited twice)
	Troisi et al (2006) (b) [Troisi et al's 2006a data]
	Troisi et al (2006) (c) [Troisi et al's 2006a data]
	Zupan et al (2004) (Cohrane cited twice)
	Jacklin et al (1988) (same Maccoby et al (1979))
	Marcus et al (1985) (same Maccoby et al (1979))
	Jacklin et al (1984) (same Maccoby et al (1979))
	Jacklin et al (1983) (same Maccoby et al (1979))

	Abramovich & Rowe (1973) (almost identical to Abramovich et al (1974). (hand search)
	Gandy (1971) (same Gandy (1968)) (hand search)
	Mizuno et al (1968) (Japanese version of (1969) paper)
double isotope derivative assay (N=5)	Saez & Bertrand (1969) (hand search)
	Mizuno et al (1969)
	Rivarola et al (1968)
	Gandy (1968) (hand search)
	Bertrand & Saez (1968) (hand search)
Data from unhealthy women and/or children &/or pregnancy complications or healthy women and/or children not presented separately from other cases (N=6)	Carlsen & Vanky (2010)
	Jin et al (2009)
	Pardo et al (2004) (EMBASE)
	Su et al (1996)
	Simmons (1995)
	Forest et al (1980)
Single case study (N=4)	Bertalan et al (2007)
	da Silva et al (2007)
	Holt et al (2005)
	Hensleigh et al (1975)
Aborteses (N=3)	Stern et al (1975)
	Reyes et al (1974) (hand search)
	Reyes et al (1973) (hand search)
Amniotic (N=3)	Ahluwalia et al (1992) (also arteries)
	Nagamani et al (1979) (hand search)
	Caputo et al (1974)
Mean &/or SD not given	Adkins et al (2007)

(indicating a non-normal distribution of values) (N=5)	Yuguang et al (2007) (geometric mean, no SD; hand search)
	Tan et al (1998)
	Bammann et al (1980) (hand search)
	August et al (1969)
Sampled mid gestation (N=2)	Abramovich (1974)
	Ling et al (1974)
Other (N=7)	Savage (2009) (ENT, Cochrane review)
	Owens (2008) (vocal cords, Cochrane review)
	Zupan et al (2004) (cord hygiene, Cochrane review)
	Hofmeyr (1997) (cord complications, Cochrane review)
	Ghione et al (1993) (Review; not about cord T)
	Wei et al (1990) (assay methodology)
	Mitchell (1970) (Review)

S4. Conversion of testosterone to nmol/l from other units

Units	Conversion factor
pg/mL	multiply by 0.00347
ng/dL	multiply by 0.0347
ng/100ml	multiply by 0.0347
ng %	multiply by 0.0347
ng/100ml	multiply by 0.0347
mμg/100 ml	multiply by 0.0347
μg/dl	multiply by .347
ng/mL	multiply by 3.47
μg/L	multiply by 3.47
pmol/L	multiply by .001

S5a. Methodological qualities of prospective studies included (Adapted from the Scottish Intercollegiate Guidelines Network).
1= yes; 0=no. Maximum score = 20.   Scores of: 0-6 = poor methodology; 7-13 = fair; 14-20 =good.

Quality variables	Simmer et al (1972)	Forest et al (1973)	Forest et al (1974)	Abramovich et al (1974)	Dawood & Saxena (1977)	Maccoby et al (1979)
Inclusion Criteria	0	0	0	0	1	1
Exclusion Criteria	0	0	1	0	0	1
Demographics comparable	0	0	0	0	0	0
Can the number of participating centers be determined	0	0	1	0	0	1
Has the source of cord T been identified (e.g. vein)	1	1	1	1	1	1
Are the mother's baseline characteristics comparable in the two groups	0	0	1	0	0	0
Are the children's baseline characteristics comparable in the two groups	1	1	1	0	0	1
Can the number of hospital staff taking cord samples be determined	0	0	0	0	0	0
Can the reader determine how expert the sampler was	0	0	0	0	0	0
Can the reader determine how expert the lab technician (assayer) was	1	0	0	1	0	0
Is the cord sampling technique adequately described	1	0	1	1	1	1
Is the assay technique adequately described	1	1	1	1	1	1
Is there any way that they have tried to standardize the cord sampling technique	1	0	1	0	1	1
Is there any way that they have tried to standardize the assay technique	1	1	1	1	1	1
Is the delivery type identified	1	1	1	0	0	1
Is the delivery type comparable in the two groups	1	1	1	0	0	1
Do authors address whether there is any missing data	0	0	0	0	0	0
Was the study period stated	0	0	0	0	0	1
Is it clear whether all the patients asked to enter the study took part	0	0	0	0	0	1
Analysis by intention to treat	0	0	0	0	0	0
Score	**9**	**6**	**11**	**5**	**6**	**13**

S5b. Methodological qualities of prospective studies included (Adapted from the Scottish Intercollegiate Guidelines Network).
1= yes; 0=no. Maximum score = 20.   Scores of: 0-6 = poor methodology; 7-13 = fair; 14-20 =good.

Quality variables	Penny et al (1979)	Miyamoto (1981)	Furuhashi et al (1982)	Shinkawa et al (1983)	Bolton et al (1989)	Herruzo et al (1993)
Inclusion Criteria	0	0	0	0	0	1
Exclusion Criteria	1	1	0	0	0	0
Demographics comparable	0	1	0	0	0	0
Can the number of participating centers be determined	0	0	0	0	0	1
Has the source of cord T been identified (e.g. vein)	1	1	1	1	1	1
Are the mother's baseline characteristics comparable in the two groups	1	0	0	0	0	0
Are the children's baseline characteristics comparable in the two groups	0	1	1	1	1	1
Can the number of hospital staff taking cord samples be determined	0	0	0	0	0	0
Can the reader determine how expert the sampler was	0	0	0	0	0	0
Can the reader determine how expert the lab technician (assayer) was	0	1	0	1	0	1
Is the cord sampling technique adequately described	1	0	1	0	1	1
Is the assay technique adequately described	1	1	1	1	1	1
Is there any way that they have tried to standardize the cord sampling technique	0	0	1	0	1	0
Is there any way that they have tried to standardize the assay technique	1	1	1	1	1	1
Is the delivery type identified	1	1	1	1	1	1
Is the delivery type comparable in the two groups	1	1	1	0	0	1
Do authors address whether there is any missing data	0	0	0	0	0	0
Was the study period stated	0	0	0	0	0	1
Is it clear whether all the patients asked to enter the study took part	0	0	0	0	0	0
Analysis by intention to treat	0	0	0	0	0	0
Score	**8**	**9**	**8**	**6**	**7**	**11**

S5c. Methodological qualities of prospective studies included (Adapted from the Scottish Intercollegiate Guidelines Network).
1= yes; 0=no. Maximum score = 20.   Scores of: 0-6 = poor methodology; 7-13 = fair; 14-20 =good.

Quality variables	Simmons et al (1994)	Maffeis et al (1999)	Troisi et al (2003)	Van de Beek et al (2004)	Gol et al (2004)	Anderson et al (2010)
Inclusion Criteria	1	0	1	1	1	1
Exclusion Criteria	1	0	1	1	0	1
Demographics comparable	1	0	0	0	1	1
Can the number of participating centers be determined	0	0	1	1	0	0
Has the source of cord T been identified (e.g. vein)	1	1	1	1	0	1
Are the mother's baseline characteristics comparable in the two groups	1	1	1	1	1	1
Are the children's baseline characteristics comparable in the two groups	0	1	0	1	1	0
Can the number of hospital staff taking cord samples be determined	0	0	0	0	0	0
Can the reader determine how expert the sampler was	0	0	0	0	0	0
Can the reader determine how expert the lab technician (assayer) was	0	1	0	0	1	0
Is the cord sampling technique adequately described	1	1	1	0	0	1
Is the assay technique adequately described	1	1	1	1	1	1
Is there any way that they have tried to standardize the cord sampling technique	1	1	1	0	0	1
Is there any way that they have tried to standardize the assay technique	1	1	1	1	1	1
Is the delivery type identified	1	1	1	1	1	1
Is the delivery type comparable in the two groups	1	1	1	1	1	1
Do authors address whether there is any missing data	1	0	1	1	0	1
Was the study period stated	0	0	1	1	0	0
Is it clear whether all the patients asked to enter the study took part	0	0	1	0	0	0
Analysis by intention to treat	0	0	0	0	0	0
Score	**12**	**10**	**14**	**12**	**9**	**12**