UNIVERSITY OF Southampton
School of Medicine

Predictive tests: Can we know what we don't know?

Nick Macklon

Chair, Obstetrics and Gynaecology

University of Southampton

UNIVERSITY OF Southampton
School of Medicine

From diagnosis…

- Symptomatology
- Incidence
- Etiology
- Diagnosis
- Treatment

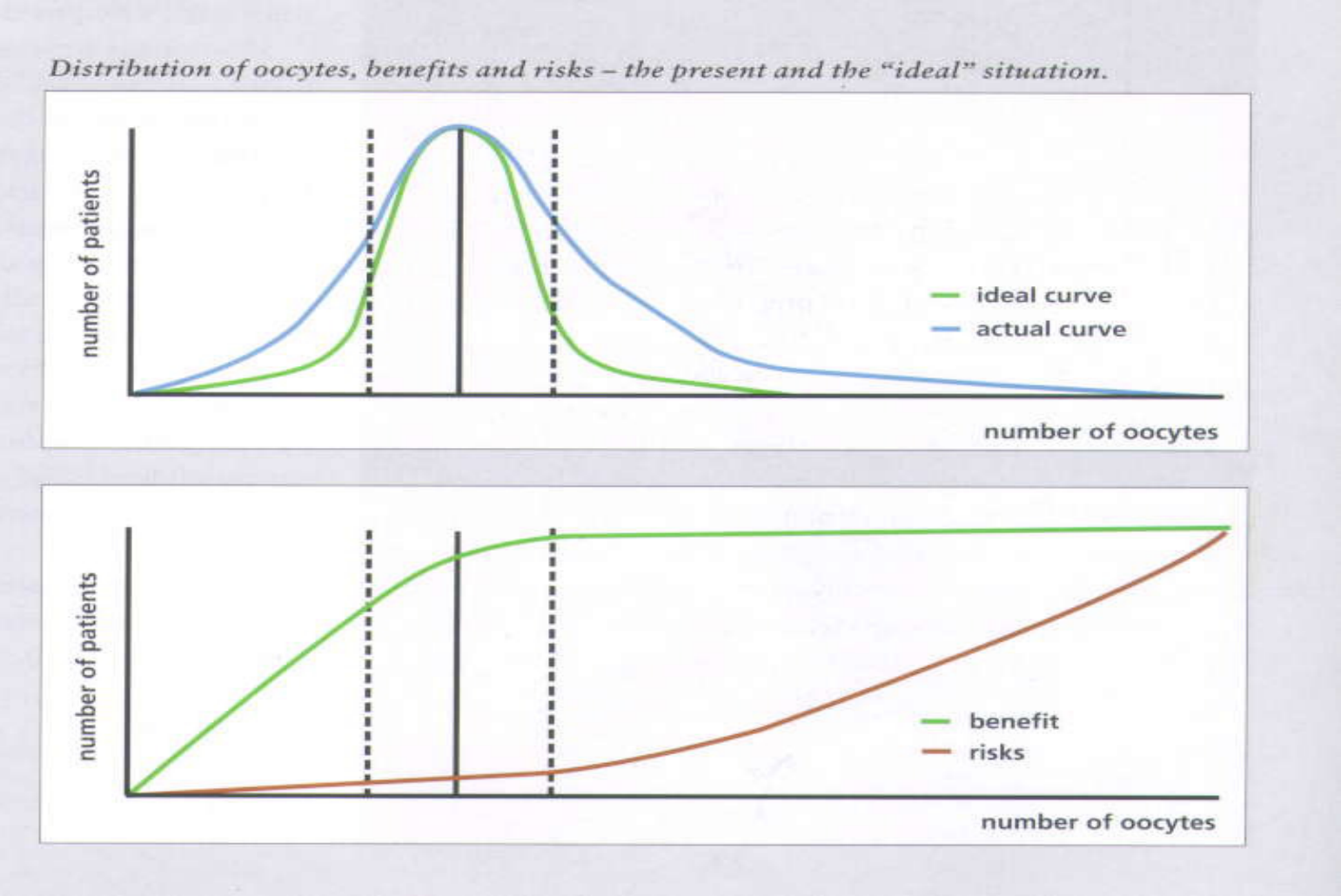
…to prognosis

UNIVERSITY OF Southampton
School of Medicine

- Timely intervention
- Avoid overtreatment
- Minimise risks
- Maximise outcomes
- Cost-efficiency aspects
- Aid decision making for patient and doctor

UNIVERSITY OF Southampton
School of Medicine

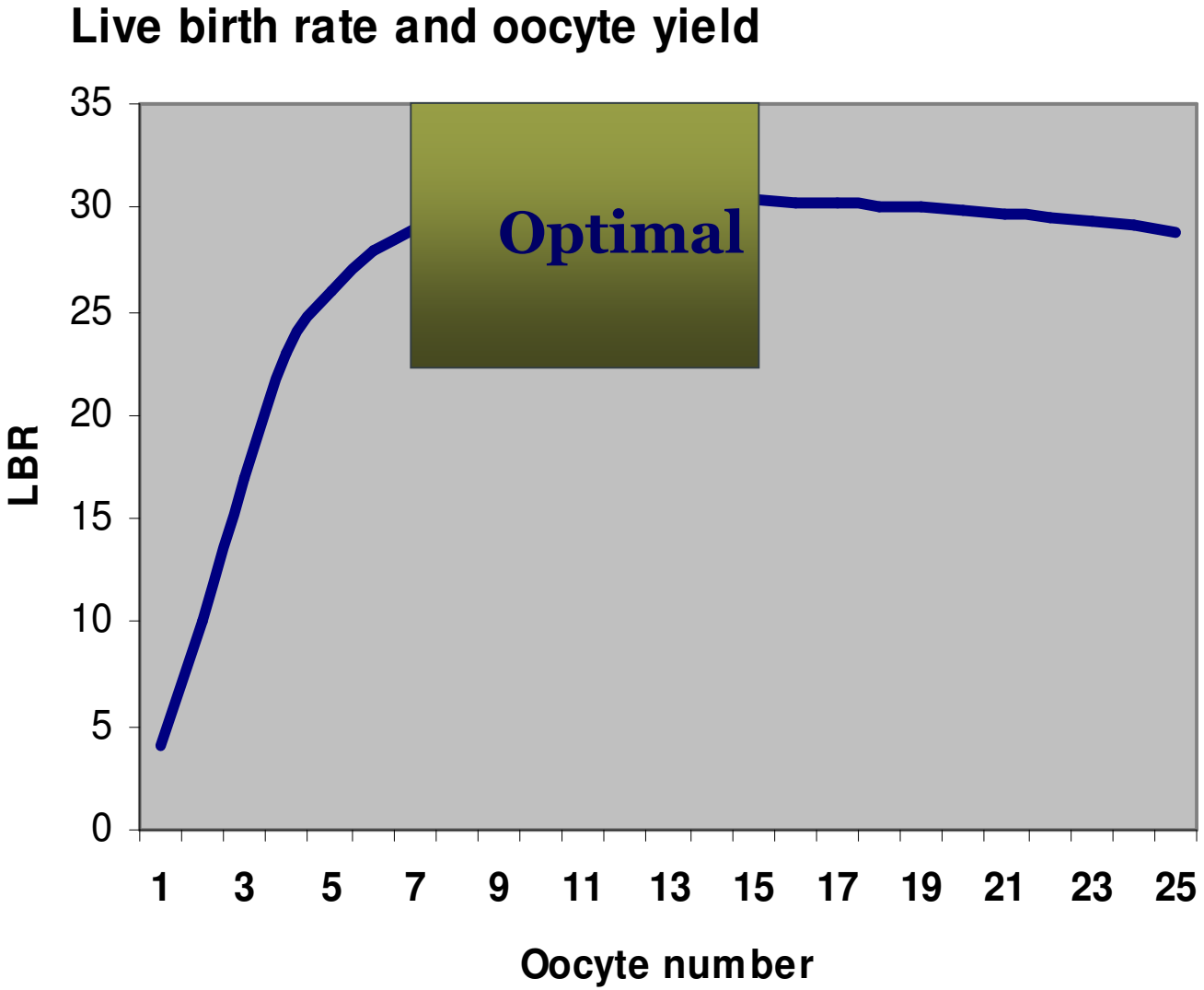
The ideal balance between risks and benefits



Popovic-Todorovic B et al, Human Reprod 2003; 18:2275-2282

Role for ovarian response prediction? Balance…

UNIVERSITY OF Southampton
ool of Medicine



Van der Gaast et al, RBM Online, 2006

UNIVERSITY OF Southampton
School of Medicine

Can we know what we don't know?

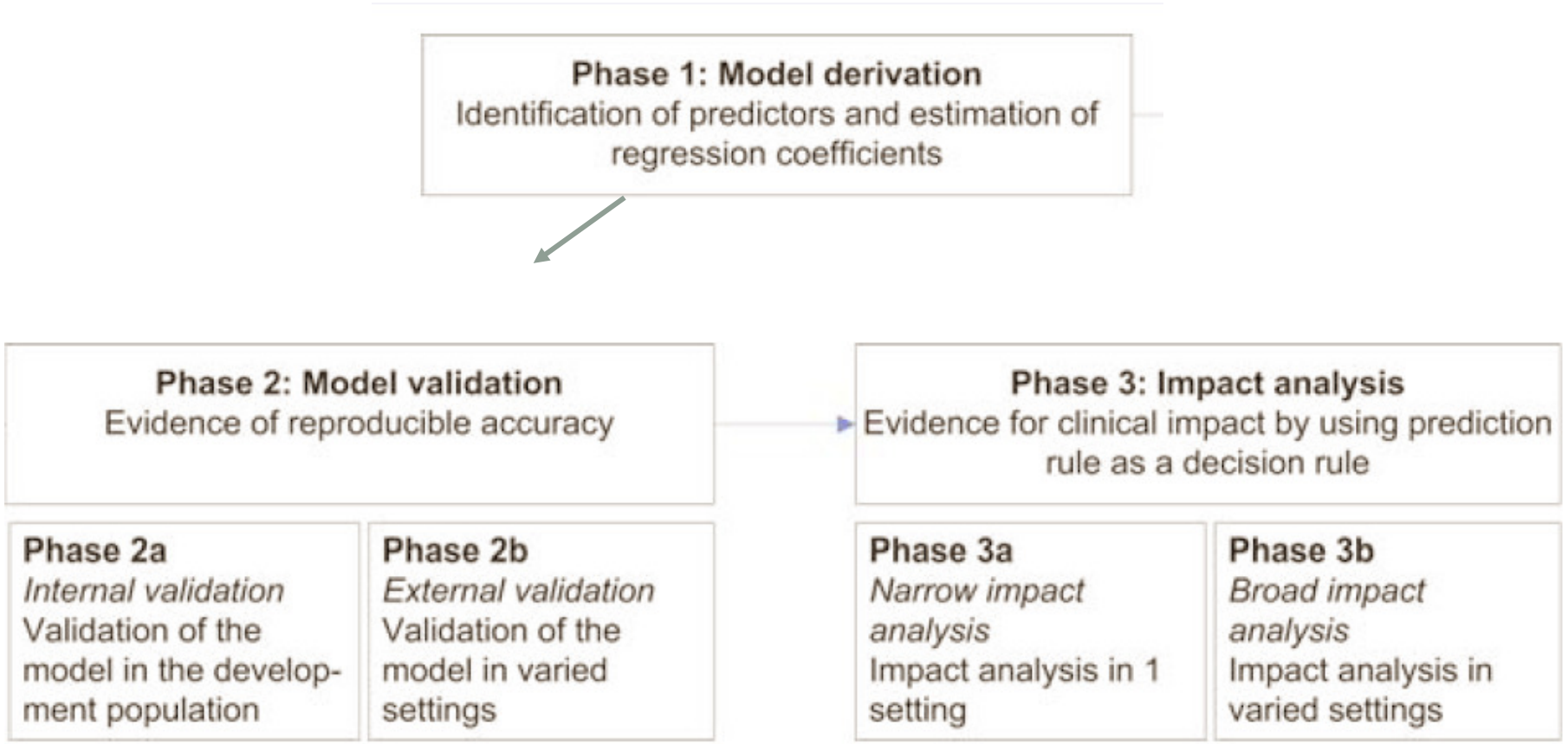
1. Building and evaluating a Prediction Model

2. Using prediction models to guide ovarian stimulation

3. Using prediction models to select patients for mild stimulation

Phases of Prediction Model development

UNIVERSITY OF Southampton
School of Medicine



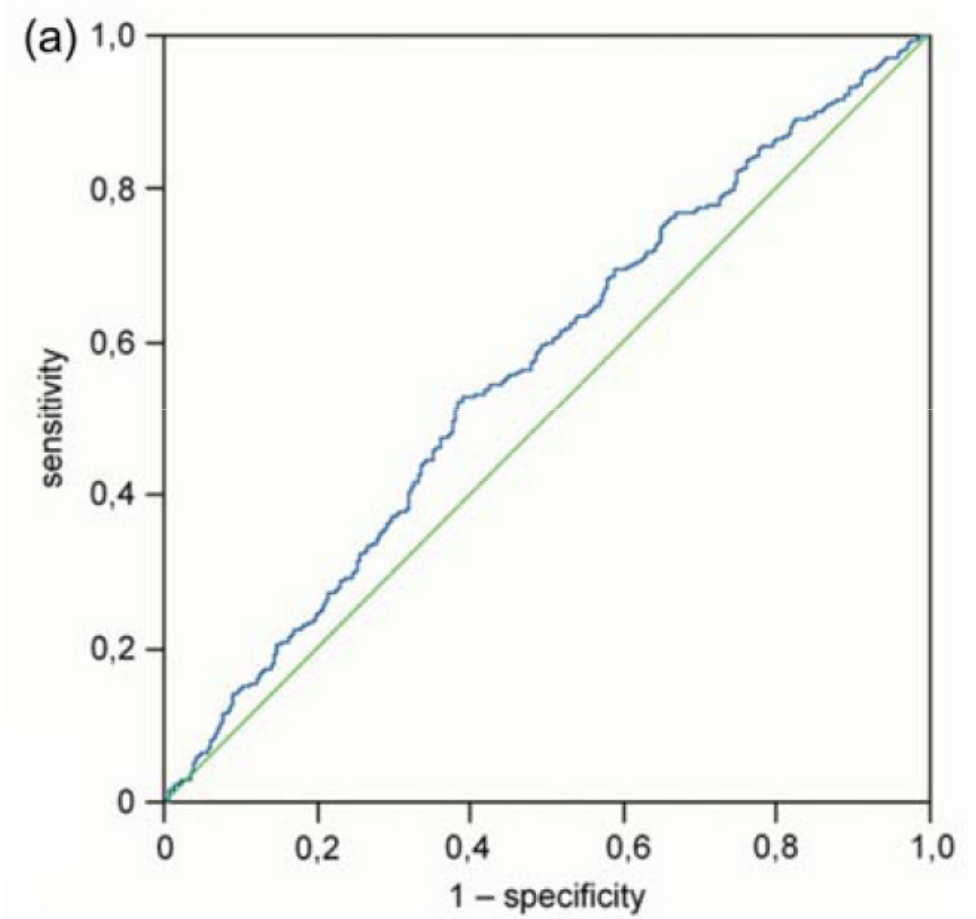
8/29 models

1/29 models

Leushuis et al, HRU 2009

Discrimination: the ability to distinguish couples who will conceive from those who will not.

UNIVERSITY OF Southampton
School of Medicine



From: Custers *et al.* External validation of model for IUI. Fertil Steril 2007.

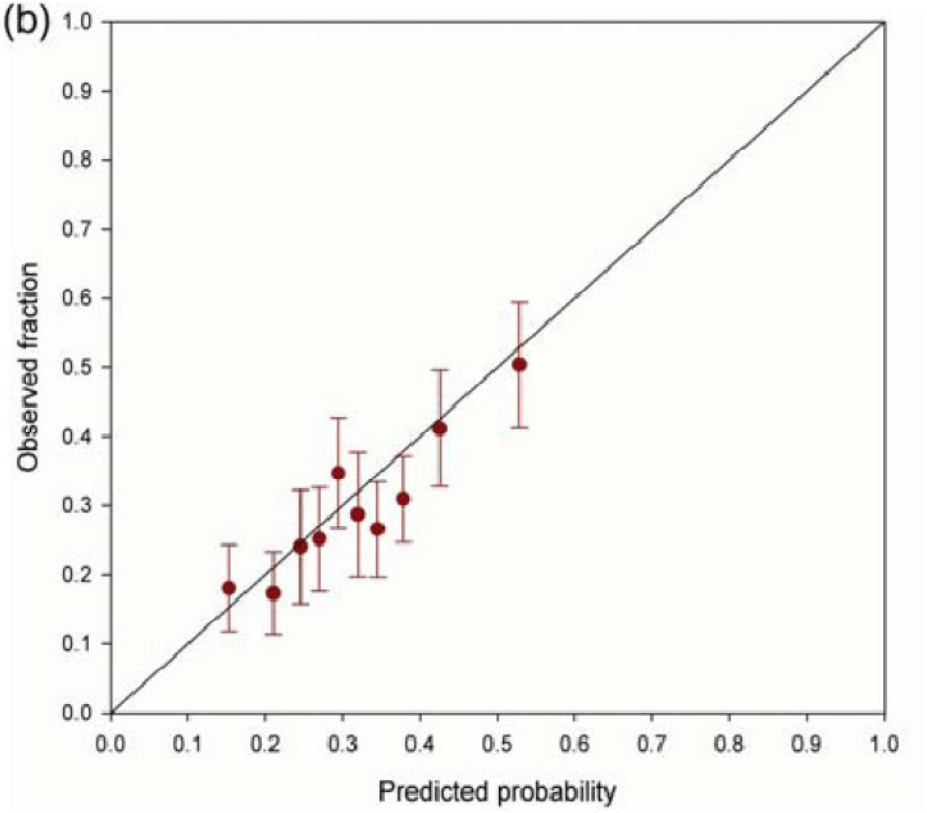
Most couples have some chance of conceiving, whereas even the most fertile couples never have a 100% chance of conception per cycle.

Consequently, discrimination will always be imperfect and to use it as a test of a model's performance is not appropriate.

Leushuis et al HR Update 2009

Calibration: the level of correspondence between the calculated pregnancy chances and the observed proportion of pregnancies

UNIVERSITY OF Southampton
School of Medicine



From: van der Steeg et *al.*. Pregnancy is predictable: a large-scale prospective external validation of the prediction of spontaneous pregnancy in subfertile couples. Human Reproduction 2007.

Well-calibrated models are able to classify individuals into **clinically useful** prognostic strata on the basis of the calculated probabilities of a pregnancy with and without treatment.

Leushuis et al HR Update 2009

Human Reproduction, Vol.1, No.1 pp. 1–5, 2009
doi:10.1093/humrep/dep109

human reproduction **OPINION**

Evaluating prediction models in reproductive medicine

S.F.P.J. Coppus[1,2,3,4], F. van der Veen[1], B.C. Opmeer[2], B.W.J. Mol[1,3], and P.M.M. Bossuyt[2]

[1]Department of Obstetrics and Gynaecology, Centre for Reproductive Medicine, Academic Medical Centre, Amsterdam, The Netherlands [2]Department of Clinical Epidemiology and Biostatistics, Academic Medical Centre, Room J1B-216-1, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands [3]Department of Obstetrics and Gynaecology, Máxima Medical Centre, Veldhoven, The Netherlands

[4]Correspondence address. E-mail: s.f.coppus@amc.uva.nl; s.coppus@mmc.nl

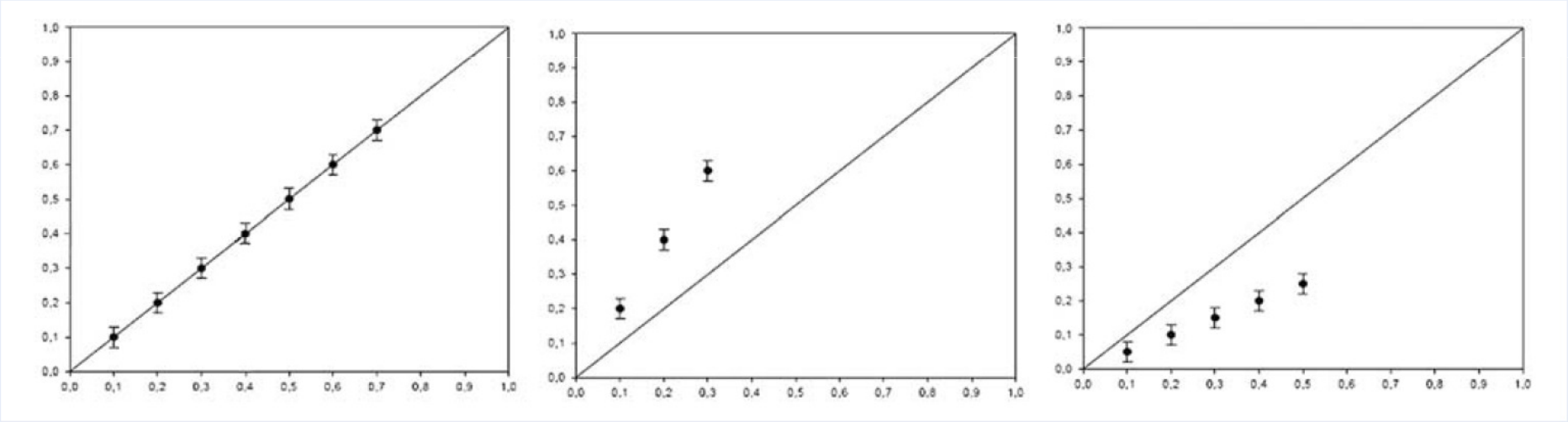


Figure 2 Calibration plots with calculated probability on the *X*-axis and observed proportion on the *Y*-axis. The left plot shows perfect calibration. The middle plot demonstrates a model that tends to give underestimated probabilities, whereas the plot on the right shows systematic overestimation.

Prediction Models: Summary

UNIVERSITY OF Southampton
School of Medicine

1. Require validation in external population

2. ROC curves: limited importance

3. In clinical practice what is more important is:

- Calibration: predicted versus observed pregnancy rates
- Clinically useful distribution of probabilities
- Ability to correctly identify appropriate form of management

UNIVERSITY OF Southampton
School of Medicine

Using Prediction Models to Guide Ovarian Stimulation

Predicting Response to Individualise Dose: The CONSORT study

UNIVERSITY OF Southampton School of Medicine

Computer model developed to predict FSH starting dose in women <35 years undergoing ART

Predictive factors in the model:

1. Basal FSH
2. Body mass index
3. Age
4. Antral follicle count

Prospective study adapting the dosage according to the model

Olivennes F, et al. Reprod Biomed Online 2009; 18(2): 195–204

CONSORT stimulation results

UNIVERSITY OF Southampton
School of Medicine

	75 IU (n=48)	112.5 IU (n=45)	150 IU (n=34)	187.5 IU (n=24)	225 IU (n=10)	All (n=161)
Total IU FSH	**1102** (672)	**1287** (447)	**1632** (341)	**2044** (276)	**2573** (552)	**1498** (648)
Days FSH	**12.5** (4.4)	**11.0** (2.9)	**10.6** (1.8)	**11.0** (1.4)	**11.5** (2.4)	**11.4** (3.1)
No. cycles cancelled (%)	**12** (25.0)	**4** (8.9)	**4** (11.8)	**2** (8.3)	**2** (20.0)	**24** (14.9)
Mean (SD) number of oocytes retrieved	**8.3** (4.5)	**9.6** (6.5)	**12.1** (6.4)	**12.7** (4.3)	**8.3** (3.8)	**10.3** (5.7)

Olivennes F, et al. Reprod Biomed Online 2009; 18(2): 195–204

Calculating an Individual FSH Dose: The Copenhagen Model

UNIVERSITY OF Southampton School of Medicine

Parameter	FSH Starting Dose
Total number of follicles	50-90 IU
Total ovarian volume	50-90 IU
Ovarian blood flow (Doppler)	0-30 IU
Age	0-20 IU
Smoking	0-20 IU
	Total starting dose

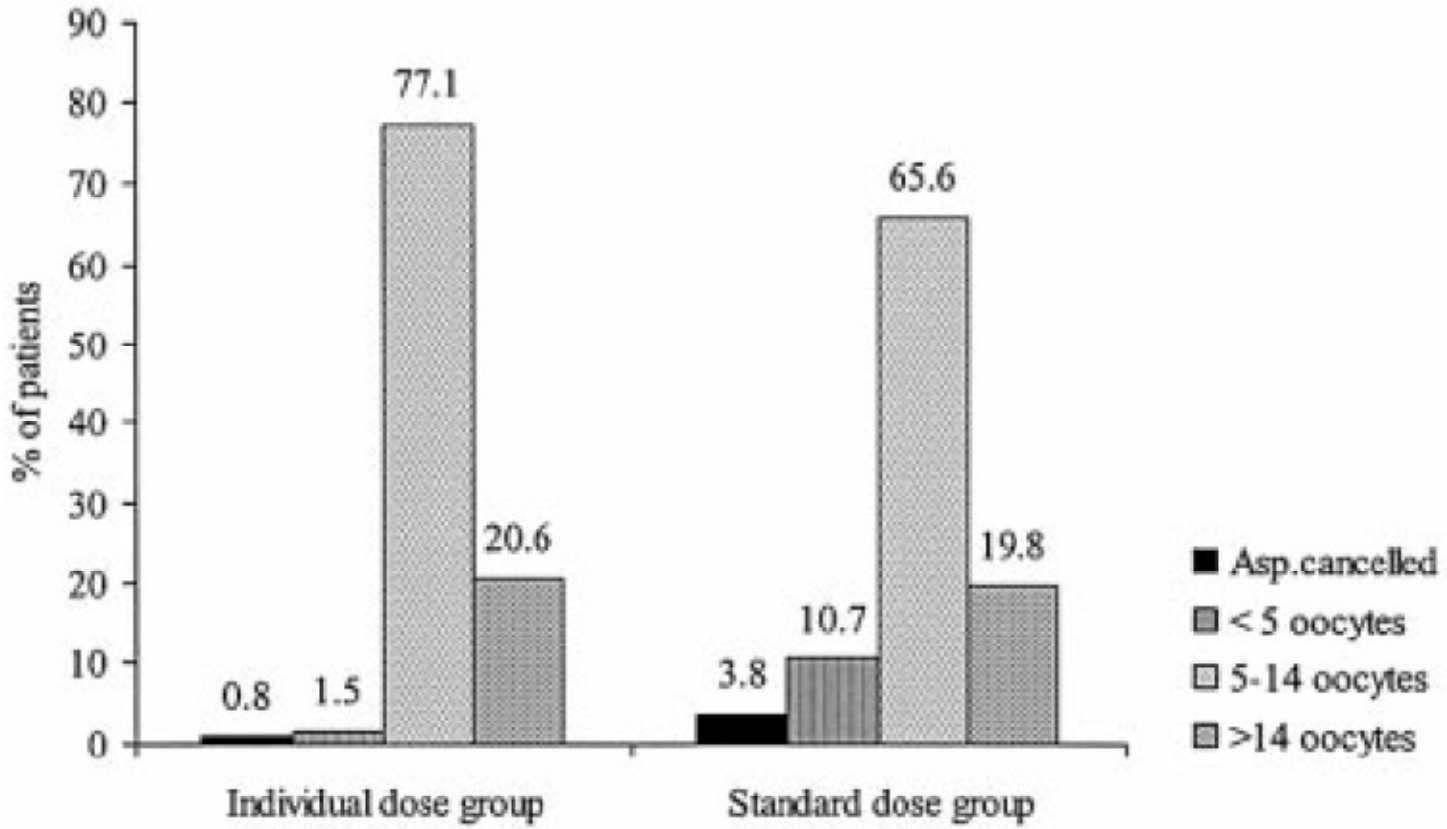
Popovic et al 2004

Human Reproduction Vol.18, No.11 pp. 2275–2282, 2003 DOI: 10.1093/humrep/deg472

UNIVERSITY OF Southampton School of Medicine

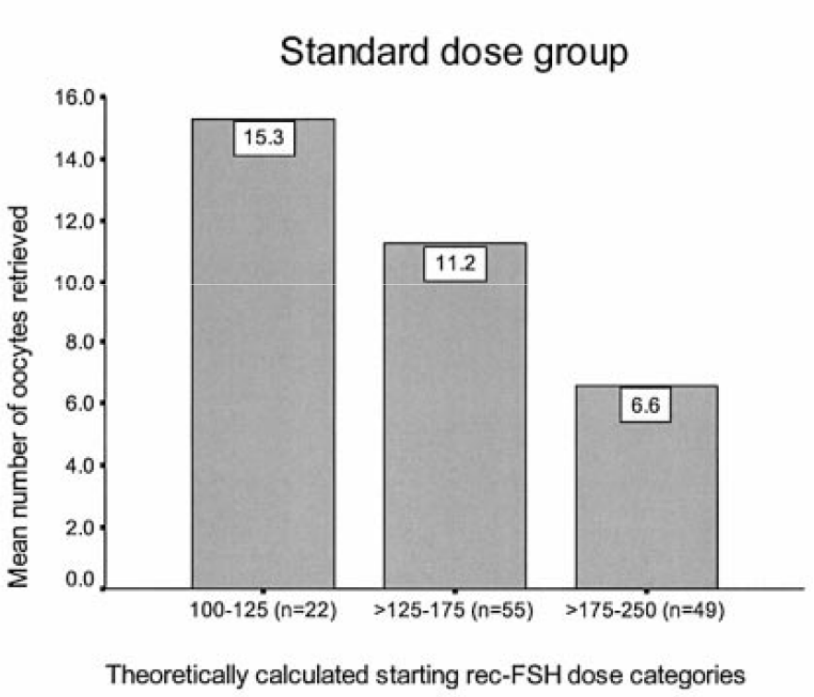
A prospective randomized clinical trial comparing an individual dose of recombinant FSH based on predictive factors versus a 'standard' dose of 150 IU/day in 'standard' patients undergoing IVF/ICSI treatment

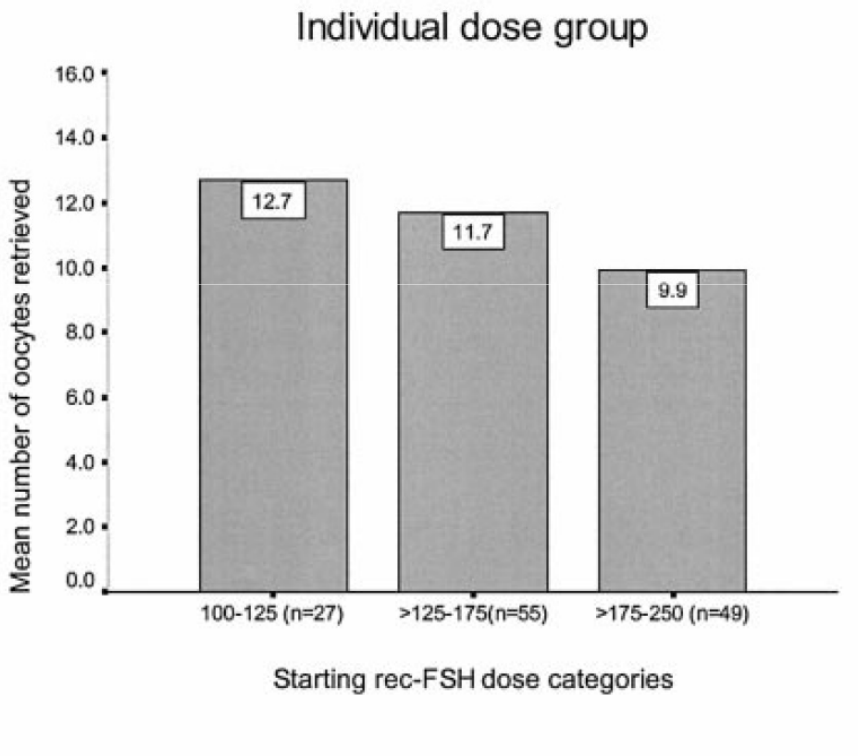
B.Popovic-Todorovic[1,3], A.Loft[1], H.Ejdrup Bredkjæer[2], S.Bangsbøll[1], I.K.Nielsen[2] and A.Nyboe Andersen[1]



Mean numbers of oocytes in relation to the arbitrarily designated starting rFSH dose categories

UNIVERSITY OF Southampton
School of Medicine

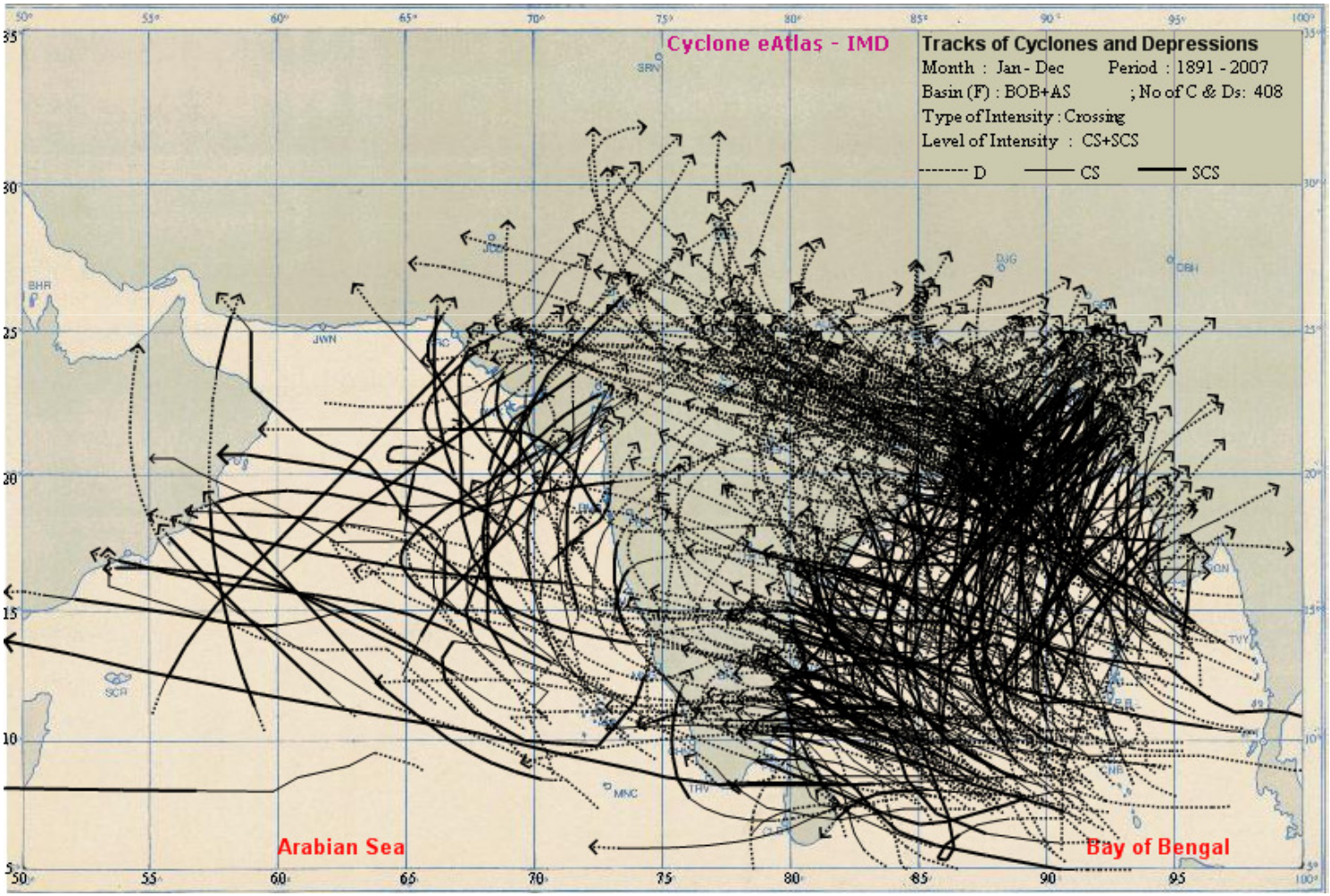




Popovic-Todorovic et al HR 2003

UNIVERSITY OF Southampton
School of Medicine

..but what about predicting PREGNANCY?

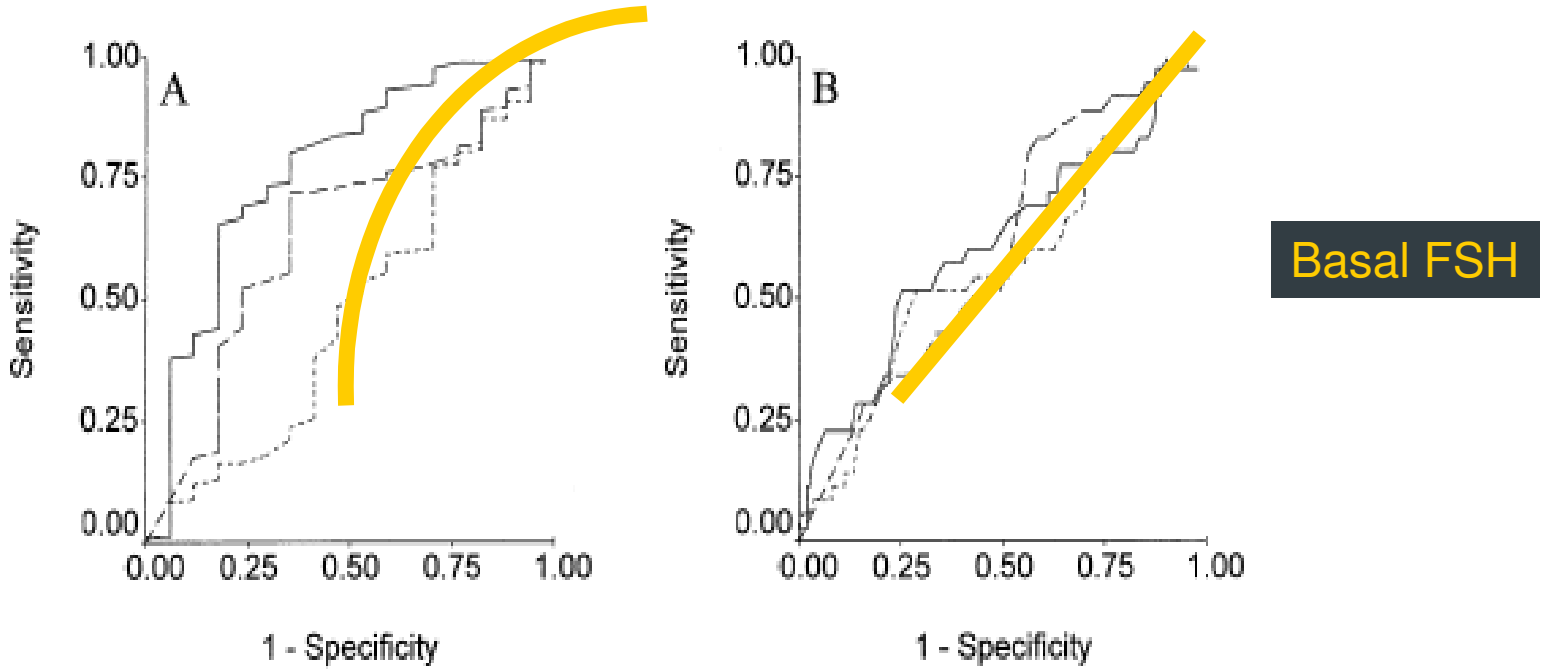


Basal FSH prediction of outcome in **antagonist** cycles

UNIVERSITY OF Southampton
School of Medicine

FIGURE 1

Receiver operating characteristic (ROC) curves and the area under the curve (AUC) illustrating the predictive value of baseline levels of FSH (—), E_2 (– – –), and LH (· · ·) in cycles from patients with normal prognosis. The levels indicated for (**A**) ovarian response and (**B**) achievement of pregnancy were FSH 0.77 ($P<.01$) and 0.61 ($P=.04$); E_2 0.63 ($P=.08$) and 0.60 ($P=.06$), and LH 0.47 ($P=.67$) and 0.57 ($P=.20$), respectively.



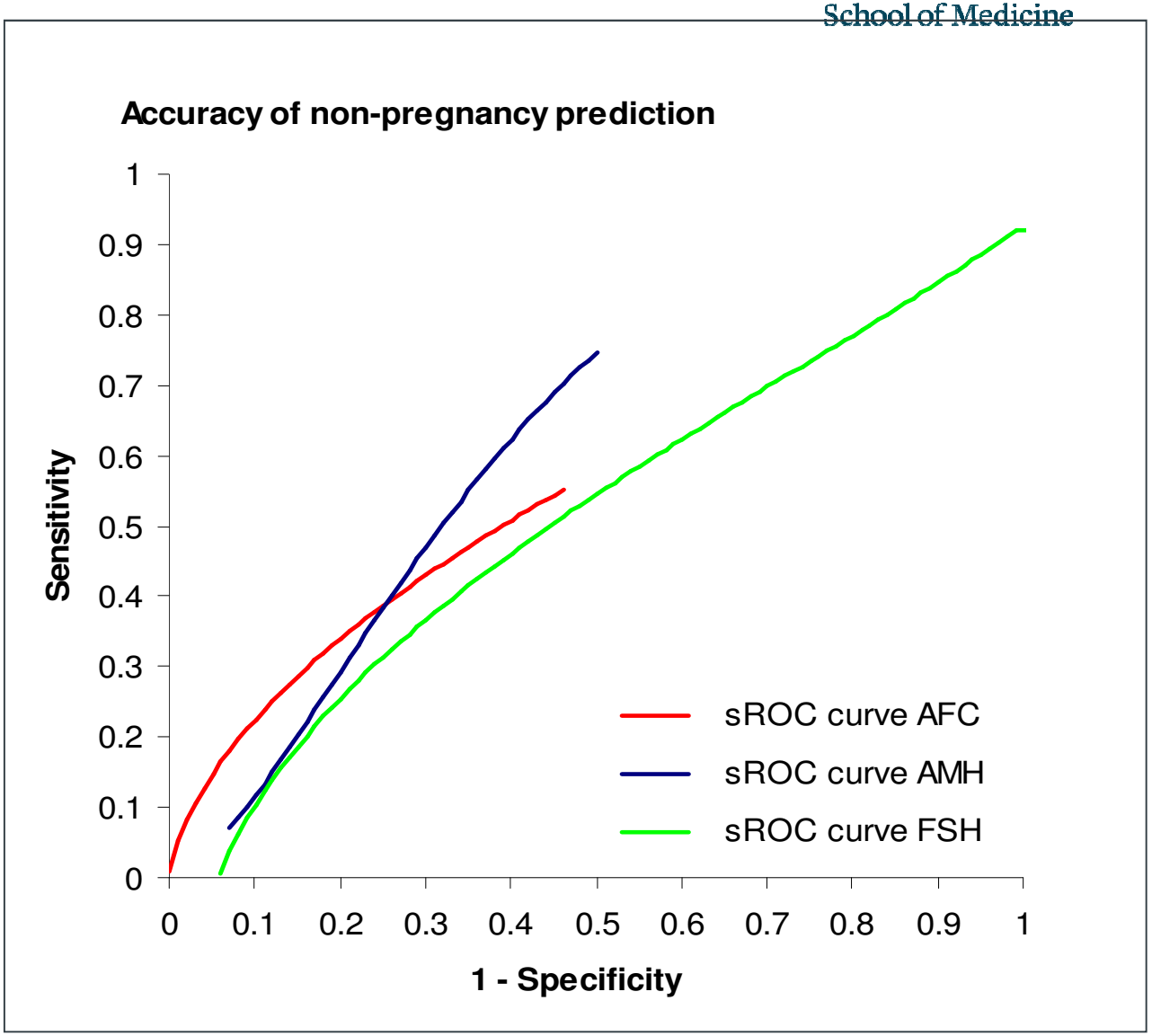
Jurema. Baseline hormones and IVF outcome. Fertil Steril 2003.

(Jurema FS, 2003)

Accuracy of the three best ORTs

UNIVERSITY OF Southampton
School of Medicine

Accuracy is poor, only at extreme cut-off levels a few zero prognosis cases may be identified

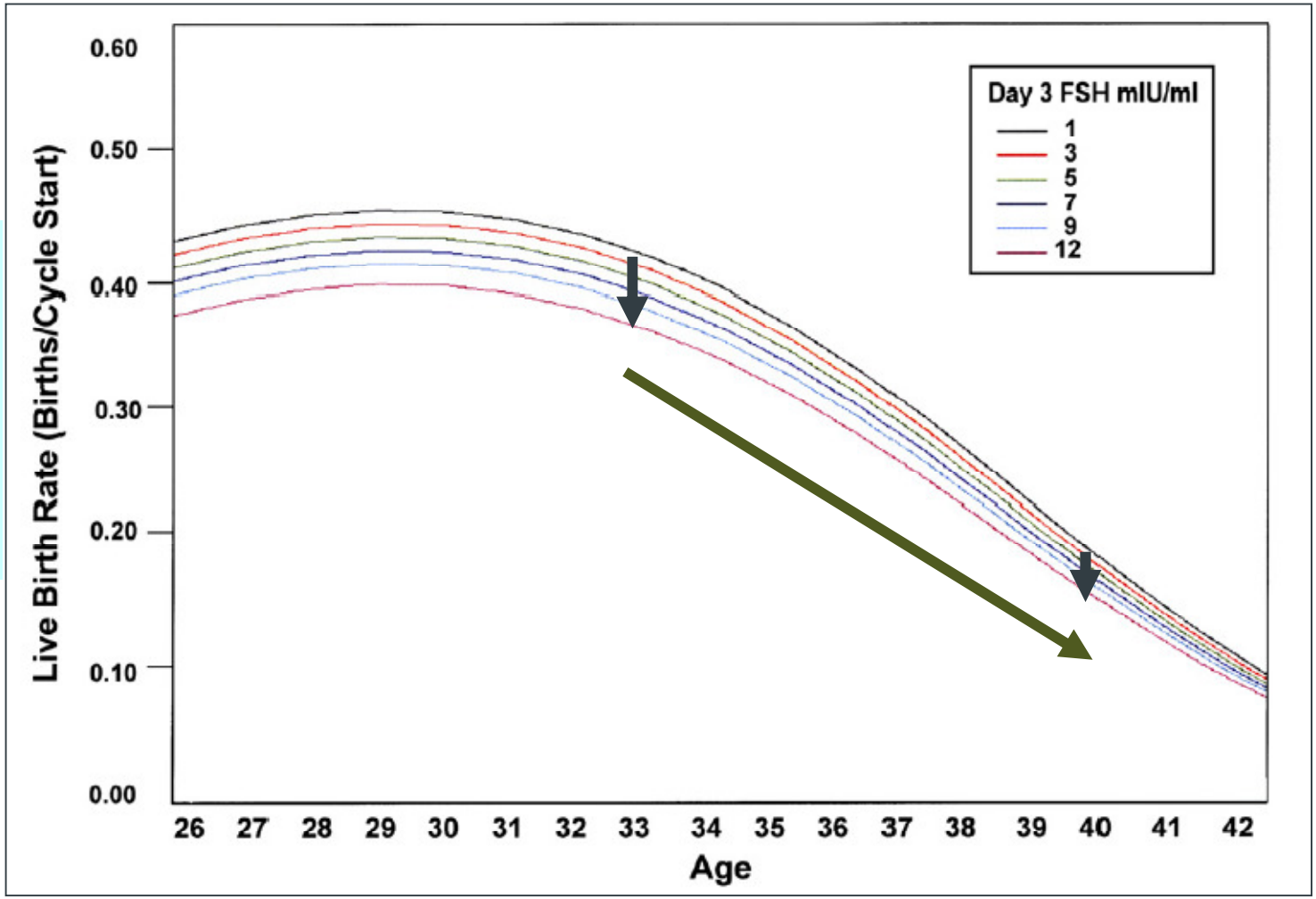


Hendriks DJ, et al. Fertil Steril 2005; 83(2): 291-301;
Broekmans FJ, et al. Hum Reprod Update 2006; 12(6): 685–718;
Broer SL, et al. Fertil Steril 2009; 91(3): 705–714

Female age or Ovarian Reserve Test to predict live birth rate?

UNIVERSITY OF Southampton
School of Medicine

The added effect of ORT to knowing age is marginal



Henne, Fertil Steril 2008;89:104–10

UNIVERSITY OF Southampton
School of Medicine

Using Prediction Models to Select Patients for Mild Stimulation

Predictors of low response to mild ovarian stimulation initiated on cycle day 5 for IVF

M.F.G. Verberg[1,3], M.J.C. Eijkemans[1,2], N.S. Macklon[1], E.M.E.W. Heijnen[1], B.C.J.M. Fauser[1] and F.J. Broekmans[1]

Table III: Multivariable analysis for cancellations due to poor response in the mild CD 5 stimulation protocol; the ability of the model measured by the area under the ROC curve was 0.69 (95% CI: 0.58–0.79)

	P-value	Odds Ratio (95% CI)[a]	Cumulative AUC
Duration of infertility	0.033	1.24 (1.02, 1.50)	0.60
Menstrual cycle length	0.034	0.75 (0.59, 0.98)	0.67
Secondary infertility	0.13	2.08 (0.82, 5.27)	0.68
BMI (Kg/m^2)	0.26	1.10 (0.93, 1.29)	0.69

Verberg et al Hum Reprod 2007

Performance of the model

UNIVERSITY OF Southampton
School of Medicine

Table IV: Clinical value of the model for cancel prediction with test characteristics at several probability cut-offs

Cut-off value for the probability of cancel	0.10	0.15	0.20	0.25	0.30
Sensitivity	87	77	43	37	33
Specificity	29	54	74	84	92
PPV	21	26	27	33	48
NPV	91	92	86	89	87
% of patients that will change protocol	89	62	29	19	12
Number of cancels unpredicted (*n* (%))	4 (13%)	7 (23%)	17 (57%)	19 (63%)	20 (67%

- Model predicts 33% of cancellations with 8% false positive rate.

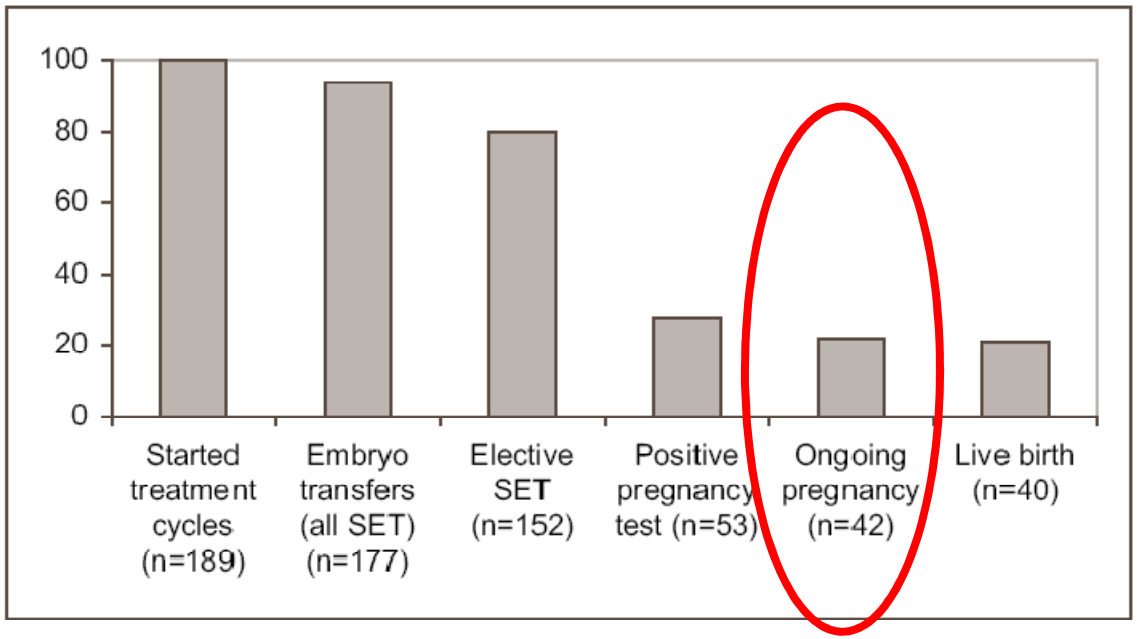
- Results in similar cancellation rate to that observed in standard GnRH antagonist protocol

Predictors of ongoing pregnancy after single-embryo transfer following mild ovarian stimulation for IVF

Marieke F. G. Verberg, M.D.,[a] *Marinus J. C. Eijkemans, Ph.D.,*[b] *Nicholas S. Macklon, M.D., Ph.D.,*[a] *Esther M. E. W. Heijnen, M.D., Ph.D.,*[a] *Bart C. J. M. Fauser, M.D., Ph.D.,*[a] *and Frank J. Broekmans, M.D., Ph.D.*[a]

FIGURE 1

Cumulative allocation of included patients.
SET = single-embryo transfer.



Verberg. Prediction of ongoing pregnancy after SET. Fertil Steril 2008.

Verberg et al FS 2008

Multivariate analysis: predictors of pregnancy

UNIVERSITY OF Southampton School of Medicine

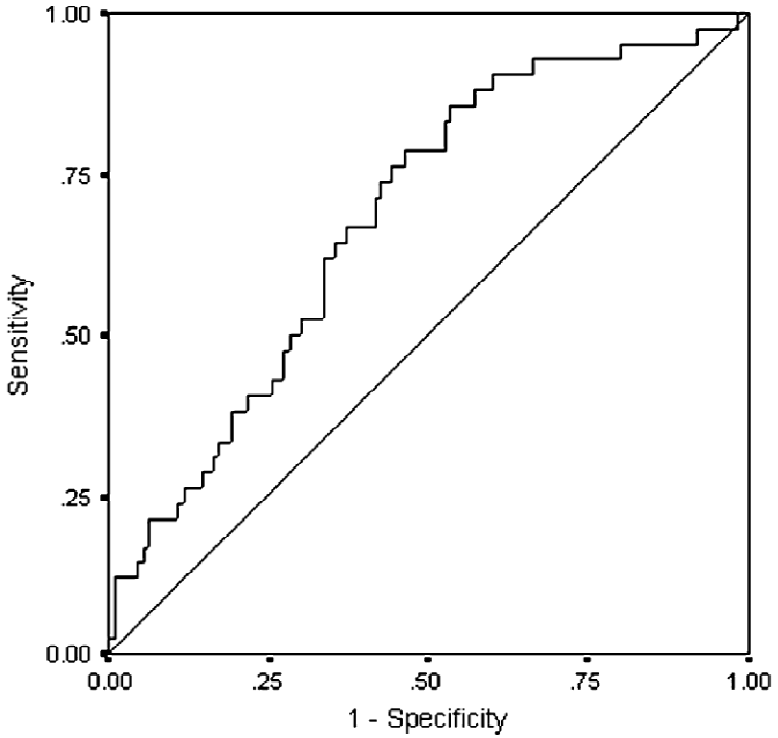
	Odds ratio (95% confidence interval)[a]	Cumulative AUC	*P* value
Body mass index (BMI)	0.89 (0.76, 1.03)	0.59	.108
Total amount of rFSH used[b]	0.92 (0.83, 1.03)	0.63	.146
Number of oocytes	0.93 (0.85, 1.01)	0.67	.077
Top-quality embryo availability	2.18 (0.93, 5.09)	0.68	.072

Performance of Model

UNIVERSITY OF Southampton
School of Medicine

FIGURE 2

Receiver operating characteristic (ROC) curve of the prediction model for the occurrence of ongoing pregnancy after elective single-embryo transfer following mild ovarian stimulation for IVF. The area under the final ROC curve is 0.68.



Verberg. Prediction of ongoing pregnancy after SET. Fertil Steril 2008.

Low area under ROC but:

Using this model-

By transferring 2 embryos in women with <20% chance of ongoing pregnancy:

→

Pregnancy rate 14% 26%

→

Multiple rate 0% 2%

Table VI Overview of the parameters of the prediction models for pregnancy after IVF (expressed as HRs or ORs)

	van Weert *et al.* (2008)	Lintsen *et al.* (2007)	Verberg *et al.* (2007)	Carrera *et al.* (2007)	Ottoson *et al.* (2007)	Ferlitsch *et al.* (2004)	Hunault *et al.* (2002)	Bancsi *et al.* (2000)	Stolwijk *et al.* (2000)	Minaretzis *et al.* (1998)	Commenges-Duces *et al.* (1998)	Templeton *et al.* (1996)	Stolwijk *et al.* (1996)[l]	Bouckaert *et al.* (1994)	Haan *et al.* (1991)	Hughes *et al.* (1989)	Nayudu *et al.* (1989)	Presence of the parameter in the prediction model (number out of 17 models)
Type of analysis	LR	CR	LR	LR	LR	LR	LR	LR	CR	LR	LR	LR	LR	LR	LR	LR	LR	
Couple factors																		
Duration of subfertility		0.97													0,64			2
Secondary subfertility	1.4	1.11							1.34[g]									3
Previous succesful IVF											2.12							1
Previous unsuccessful IVF																194[q]		1
Female factors																		
Female age	0.94	[a]		0.89	0.74[d]		0.98	0.95	1.73[h]	0.93	0.28[j]	1.01[k]	0.94	2.05[m]	0.56[p]	1.1[r]		14
Body mass index			0.89		0.88[e]	0.84												3
Unexplained subfertility															1.5			1
Basal FSH					0.55	0.77		0.90										3
Tubal reasons for IVF	0.4											0.93			0.65			3
Tuboperitoneal disease								0.24										1
Endometriosis		1.05[b]																1
Cervical factor subfertility		1.04[b]																1
Previous IVF live birth												2.14						1
Previous IVF preg., no live birth												1.35						1
Previous live birth (no IVF)												1.26						1
Previous preg.(no IVF), no live birth												1.12						1
≥1 previous pregnancy													2.26					1
History of unsuccessful IUI	0.59																	1
Cycle number	1.4																	1
Total amount of rFSH used			0.92[c]															1
Number of ampoules											0.98							1
Antral follicle count				1.15														1
Estradiol stimulation Day 4				1.01														1
hCG																	1.06	1
Pregnancy type follicle																	62[q]	1
Total protein																	10301[q]	1
E_2FD (first day E_2 increase)																	4.5	1
Male factors																		
Sperm motility (mean%)	0.98																	1
Sperm morphology (mean%)	1.01																	1

Smoking and Pregnancy rate after IVF

UNIVERSITY OF Southampton
School of Medicine

Waylen et al, HRU 2009

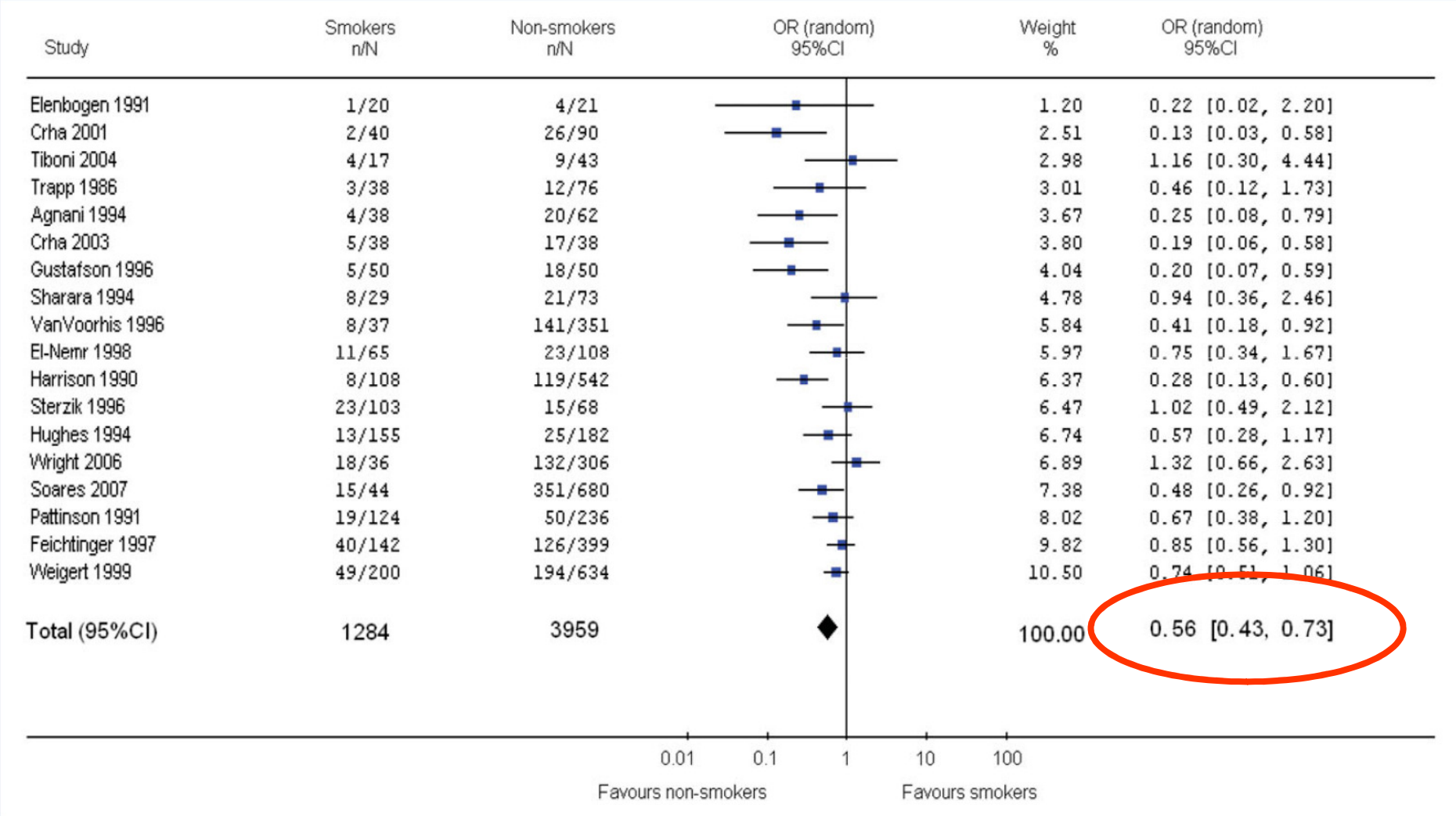


Figure 2 Odds ratio of clinical pregnancy rate per cycle.

Total events: 236 (smokers), 1303 (non-smokers). Test for heterogeneity: $\chi^2 = 33.27$, df = 17 ($P = 0.01$), $I^2 = 48.9\%$. Test for overall effect: $z = 4.26$ ($P < 0.0001$).

Conclusions 1

UNIVERSITY OF Southampton
School of Medicine

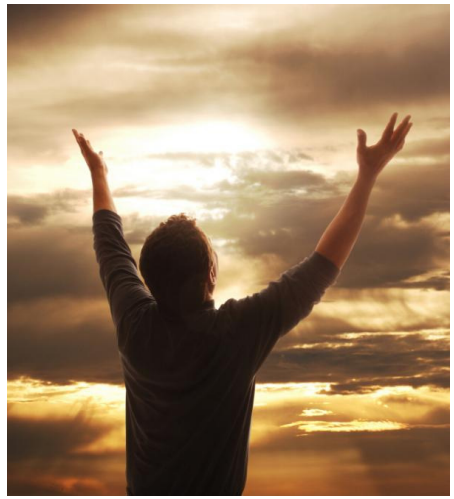
- Doctors are becoming ‘Prognosticians’

- Tests may have poor discrimination but still be useful.
- Prediciting ovarian response easier than predicting pregnancy
- Prognostic factors indicate therapeutic opportunities

- The tests and models that serve us are imperfect but for can improve some outcomes.

Predictive tests:
Can we know what we don't know?

UNIVERSITY OF Southampton
School of Medicine



- There are known knowns.
- These are things we know that we know.
- There are known unknowns.
- That is to say, there are things that we know we don't know.
- But there are also unknown unknowns.
- There are things we don't know we don't know.

(Rumsfeld et al, 2002)