Reply: Clinical trial registry alone is not adequate: On the perception of possible end point switching and *P*-hacking

Sir,

We are grateful for the opportunity to respond to the letter written by Hill, Connell and Patounakis. We can reassure the authors that their impression of 'end point switching and *P*-hacking' is incorrect. The primary outcome was confirmed *a priori* with our trial statistician (L.H.) before the start of the trial and analysis independently carried out by him following conclusion of the trial. The sample size for this trial was determined by a power calculation based on this primary outcome and relevant data from our previous trial (Abbara *et al.*, 2015). The primary outcome was thus pre-defined and recorded in the statistical analysis plan (SAP), which was reviewed by Human Reproduction editorial team and reviewers during the submission process.

Hill et al. request further clarification for the choice of primary outcome used in this trial. As stated in the introduction and discussion of the manuscript (Abbara *et al.*, 2017), we deliberately chose a patient-centric primary outcome to address the variability in response observed in our previous trial using a single bolus of kisspeptin (Abbara *et al.*, 2015). From a patient's perspective, the chance of achieving a clinically effective outcome is a more meaningful outcome than the average oocyte yield across a group, for which the variability in response could result in either a poor or good response being encountered for a particular patient. A number of trials in high impact journals have chosen similar primary outcomes assessing the proportion of patients who meet a threshold for efficacy, rather than presenting only the mean difference across a group (Garg *et al.*, 2017). In the UK, there is a drive to encourage the choice of primary outcomes that are meaningful to patients rather than only to researchers (termed 'patient and

public involvement in research'; PPI), with patients viewed as partners in the design and choice of outcomes for clinical research trials rather than merely as participants (Bagley *et al.*, 2016). Hence, the choice of primary outcome for this trial was made following discussions with patients from our previous trials (Jayasena *et al.*, 2014; Abbara *et al.*, 2015).

We agree with Hill and colleagues that it is important to additionally present traditional markers of oocyte maturation, such as the number of mature oocytes and oocyte maturation rate (as presented in Table 2 of the manuscript) (Abbara *et al.*, 2017), however it is important to recognize that these markers also have limitations as measures of trigger efficacy. The number of mature oocytes retrieved heavily depends on the number of appropriately sized follicles on the day of trigger. In 2007, Shapiro and colleagues conducted a retrospective analysis comparing the efficacy of hCG and GnRH agonist to trigger oocyte maturation (Shapiro *et al.*, 2007). They observed that GnRH agonist resulted in a significantly higher number of oocytes retrieved (28.8) when compared to hCG (21.6) (Shapiro *et al.*, 2007). However in this retrospective study, patients receiving GnRH agonist had a greater number of follicles on the day of trigger (GnRH agonist 34.2 follicles; hCG 21.7 follicles) making it difficult to accurately compare trigger efficacy between the two groups (Shapiro *et al.*, 2007). Thus in their later work, Shapiro introduced the concept of an 'oocyte yield', whereby the number of oocytes collected is corrected for the number of follicles on the day of trigger (Shapiro *et al.*, 2011). They reported a mature oocyte yield (mature oocytes as a proportion of follicles of ≥10mm on the day of trigger) of 63% after GnRH agonist trigger (Shapiro *et al.*, 2011). Chen et al. used a similar approach reporting oocyte yield (oocytes as a proportion of follicles ≥10mm on the day of oocyte retrieval) and determined an oocyte yield of 61% following GnRH agonist trigger (Chen *et al.*, 2012). The threshold for the denominator of oocyte yield of 10mm follicles on the day of trigger was not derived from a strong scientific evidence-base and achieving a mature oocyte yield of 63%

following an effective and established dose of GnRH agonist suggests that the denominator may be too broad (perhaps including smaller follicles that are less likely to yield a mature oocyte). Thus, we drew upon the expert opinion of senior experienced IVF clinicians within our department, and the published literature (Blazquez *et al.*, 2014; Haas *et al.*, 2014), who considered that 14mm on the day of trigger was a follicle size threshold from which they would expect a good chance of yielding a mature oocyte if effective triggering is provided (Jayasena *et al.*, 2014; Abbara *et al.*, 2015). From our previous trials, mature oocyte yield using this denominator performed well as a measure of trigger efficacy demonstrating a reasonable dose-response (mature oocyte yield 53% at 3.2nmol/kg, 86% at 6.4-9.6nmol/kg, 121% at 12.8nmol/kg) (Abbara *et al.*, 2015). We have recently completed some work to provide a more robust evidence-base for the quantification of trigger efficacy for use in our future trials. The threshold of 60% oocyte yield was chosen *a priori* based on data from our previous trial to represent a level at which there is a good chance of progression to the latter stages of IVF treatment (Abbara *et al.*, 2015), but data from the current trial was used to demonstrate this being more relevant to the data being presented (Abbara *et al.*, 2017).

The authors also comment that the oocyte maturation rate is a more reliable measure of trigger efficacy than the mature oocyte yield. Oocyte maturation rate refers to the proportion of oocytes retrieved which are mature. The oocyte maturation rate has an inherent limitation as a measure of the trigger efficacy, as oocytes which are immature are additionally less likely to be retrieved at all, thus impacting on both the denominator as well as the numerator. In 2011, Shapiro observed that patients with insufficient serum LH levels following GnRH agonist triggering had lower mature oocyte yields, but similar oocyte maturation rates (Shapiro *et al.*, 2011). Similarly, Chen et al. observed that mature oocyte yield performed better as a measure of trigger efficacy than the oocyte maturation rate (Chen *et al.*, 2012). Several other studies, including our own, have

similarly described a lack of clear dose-response with oocyte maturation rates (Loumaye *et al.*, 2001; Levy *et al.*, 2013; Jayasena *et al.*, 2014; Abbara *et al.*, 2017).

Our group have carried out a programme of work investigating the potential of kisspeptin as a novel trigger of oocyte maturation and this was first registered on clinicaltrials.gov in 2012. The first trial was a 'proof of concept' study that aimed to assess whether kisspeptin could induce oocyte maturation at all and thus the presence of mature oocytes was chosen as the primary outcome for that trial (Jayasena *et al.*, 2014). The second trial aimed to demonstrate the efficacy and safety of kisspeptin in a population at high risk of ovarian hyperstimulation syndrome (OHSS), and thus mature oocyte yield was chosen as the primary outcome for the reasons outlined above (Abbara *et al.*, 2015). The primary outcome for the current trial reflected the aim of this trial to address the heterogeneity in response observed in our previous trial and to determine whether a second dose of kisspeptin could reduce this heterogeneity (Abbara *et al.*, 2017). We agree that the listing of the primary outcome as 'oocyte maturation' on clinicaltrials.gov lacked sufficient specificity for this trial and this should have been updated. However, as highlighted above the primary outcome in the current study was agreed with our trial statistician and was documented in the SAP prior to the start of the trial.

Hill et al. question how a 26% absolute difference can be statistically significant when the power calculation for the study was based on a 37% absolute difference. A power calculation merely provides a measure of the probability of avoiding a type 2 error with a predicted effect size, but this does not preclude that a smaller effect size will be statistically significant following subsequent data analysis. The authors suggest that exact logistic regression would have been a preferable analytical approach for our primary endpoint. However, exact logistic regression is a methodology specifically designed for use in small samples and especially in samples with sparse or zero event rates (Mehta and Patel, 1995). Mehta et al. recommend that the usual

maximum-likelihood based logistic regression be used except in cases where the data are sparse or unbalanced (Mehta and Patel, 1995). We are therefore confident that our primary analysis methodology is appropriate. Hill et. al. also suggest that a primary endpoint analysis should be performed using the same methodology as used in the power calculation. However, there is no convention in clinical trial analysis which dictates that this is necessary; neither the ICH E9 Guideline on Statistical Principles for Clinical Trials (Lewis, 1999), nor the CONSORT statement for clinical trial reporting (Schulz *et al.*, 2010) suggest that such an approach is required or even desirable. Finally, Hill et al. comment that our manuscript contains interesting data that merits publication regardless of the *P* value. We agree that the notion that a trial is either positive or negative based on a *P* value threshold of 0.05 is "overly simplistic" and "*P* values should be interpreted as a continuum wherein the smaller the *P* value, the greater the strength of the evidence for a real treatment effect" (Pocock and Stone, 2016). Thus, the marginal difference in *P* value generated through different methodology should not 'certainly be enough to change the final conclusion of the paper'.

Ali Abbara[1], Sophie Clarke[1], Rumana Islam[2], Julia K Prague[1], Alexander Comninos[1], Shakunthala Narayanaswamy[1], Deborah A Papadopoulou[1], Rachel E Roberts[1], Chioma N Izzi-Engbeaya[1], Risheka Ratnasabapathy[1], Alexander Nesbitt[1], Sunitha Vimalesvaran[1], Rehan Salim[2], Stuart A Lavery[2], Stephen R Bloom[1], Les Huson[3], Geoffrey H Trew[2], Waljit S Dhillo[1*]

[1]Department of Investigative Medicine, Imperial College London, Hammersmith Hospital, London, W12 0NN, UK. [2]IVF Unit, Hammersmith Hospital, London, W12 0HS, UK.

[3]Division of Experimental Medicine, Hammersmith Hospital, London, W12 0NN, UK.

***Correspondence address:** Prof. Waljit S. Dhillo, Department of Investigative Medicine, Imperial College London, 6th Floor, Commonwealth Building, Hammersmith Hospital, Du Cane Road, London, W12 ONN, UK. Tel: +44 208 383 3242 Fax: +44 208 383 3142 Email w.dhillo@imperial.ac.uk

References:

Abbara A, Clarke S, Islam R, Prague JK, Comninos AN, Narayanaswamy S, Papadopoulou D, Roberts R, Izzi-Engbeaya C, Ratnasabapathy R, *et al.* A second dose of kisspeptin-54 improves oocyte maturation in women at high risk of ovarian hyperstimulation syndrome: a Phase 2 randomized controlled trial. *Hum Reprod* 2017;**32**:1915–1924.

Abbara A, Jayasena CN, Christopoulos G, Narayanaswamy S, Izzi-Engbeaya C, Nijher GMK, Comninos AN, Peters D, Buckley A, Ratnasabapathy R, *et al.* Efficacy of kisspeptin-54 to trigger Oocyte maturation in women at high risk of ovarian hyperstimulation syndrome (OHSS) during in vitro fertilization (IVF) therapy. *J Clin Endocrinol Metab* 2015;**100**:3322–3331. Endocrine Society.

Bagley HJ, Short H, Harman NL, Hickey HR, Gamble CL, Woolfall K, Young B, Williamson PR. A patient and public involvement (PPI) toolkit for meaningful and flexible involvement in clinical trials – a work in progress. *Res Involv Engagem* 2016;**2**:15.

Blazquez A, Guillén JJ, Colomé C, Coll O, Vassena R, Vernaeve V. Empty follicle syndrome prevalence and management in oocyte donors. *Hum Reprod* 2014;**29**:2221–2227.

Chen SL, Ye DS, Chen X, Yang XH, Zheng HY, Tang Y, He YX, Guo W. Circulating luteinizing hormone level after triggering oocyte maturation with GnRH agonist may predict oocyte yield in flexible GnRH antagonist protocol. *Hum Reprod* 2012; 27 (5): 1351-1356

Garg SK, Henry RR, Banks P, Buse JB, Davies MJ, Fulcher GR, Pozzilli P, Gesty-Palmer D, Lapuerta P, Simó R, *et al.* Effects of Sotagliflozin Added to Insulin in Patients with Type 1 Diabetes. *N Engl J Med* 2017; DOI: 10.1056/NEJMoa1708337

Haas J, Zilberberg E, Dar S, Kedem A, Machtinger R, Orvieto R. Co-administration of GnRH-agonist and hCG for final oocyte maturation (double trigger) in patients with low number of oocytes retrieved per number of preovulatory follicles--a preliminary report. *J Ovarian Res* 2014;**7**:77.

Jayasena CN, Abbara A, Comninos AN, Nijher GMK, Christopoulos G, Narayanaswamy S, Izzi-Engbeaya C, Sridharan M, Mason AJ, Warwick J, *et al.* Kisspeptin-54 triggers egg maturation in women undergoing in vitro fertilization. *J Clin Invest* 2014;**124**:3667–3677.

Levy G, Hill MJ, Ramirez C, Plowden T, Pilgrim J, Howard RS, Segars JH, Csokmay J. Serum human chorionic gonadotropin levels on the day before oocyte retrieval do not correlate with oocyte maturity. *Fertil Steril* 2013;**99**:1610–1614.

Lewis JA. Statistical principles for clinical trials (ICH E9): An introductory note on an international guideline. *Stat Med* 1999;**18**:1903–1904.

Loumaye E, Engrand P, Arguinzoniz M, Piazzi A, Bologna S, Devroey P, Smith J, Andersen AN, Frydman R, Hedon B, *et al.* Recombinant human luteinizing hormone is as effective as, but safer than, urinary human chorionic gonadotropin in inducing final follicular maturation and ovulation in in vitro fertilization procedures: Results of a multicenter double-blind study. *J Clin Endocrinol Metab* 2001;**86**:2607–2618.

Mehta CR, Patel NR. Exact logistic regression: theory and examples. *Stat Med* 1995;**14**:2143–

2160.

Pocock SJ, Stone GW. The Primary Outcome Fails — What Next? *N Engl J Med* 2016;**375**:861–870.

Schulz KF, Kenneth P, Schulz F, Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement Web appendix 1 CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials 2010; 11: 32

Shapiro BS, Daneshmand ST, Garner FC, Aguirre M, Ross R. Comparison of human chorionic gonadotropin and gonadotropin-releasing hormone agonist for final oocyte maturation in oocyte donor cycles. *Fertil Steril* 2007; 88 (1): 237-239

Shapiro BS, Daneshmand ST, Restrepo H, Garner FC, Aguirre M, Hudson C, al. et. Efficacy of induced luteinizing hormone surge after "trigger" with gonadotropin-releasing hormone agonist. *Fertil Steril* 2011;**95**:826–828.