RBM ONLINE

ELSEVIER

Article

'Model' versus 'everyday' patients: can randomized controlled trial data really be applied to the clinic?

Eliyakim Hershkop [a], ***Linoy Segal*** [a], ***Ofer Fainaru*** [a,b], ***Shahar Kol*** [a,b,*]

[a] *Ruth and Bruce Rappaport Faculty of Medicine, Technion – Israel Institute of Technology, Haifa 3109601, Israel*

[b] *IVF Unit, Department of Obstetrics and Gynecology, Rambam Health Care Campus, Haifa 3109601, Israel*

Eliyakim Hershkop is a medical student in the graduating class of 2019 at The Ruth and Bruce Rappaport Faculty of Medicine, Technion – Israel Institute of Technology in Haifa, Israel. He is an Assistant Researcher at the Department of Obstetrics and Gynecology, IVF Unit, Rambam Health Care Campus. He has a Bachelor's degree (HONS) in Biology from Touro College and a Bachelor's degree (HONS) in Liberal Arts from Excelsior College.

A B S T R A C T

New drug approval requires a new drug to undergo rigorous clinical trials to determine its efficacy and safety. A drug is approved only for the population on which it was tested, i.e. those who meet the inclusion criteria of the trial. The aim of this study was to determine what percentage of 'real life' patients in our clinic meet the inclusion and exclusion criteria used in large-scale clinical trials required for drug registration in the field of assisted reproduction. All 265 consecutive patients with pertinent data treated in a tertiary centre IVF Unit during 2015 were surveyed. Their demographic and clinical parameters were compared with inclusion and exclusion criteria used in nine major clinical trials. Only 97 out of 265 (37%) patients met the consensus inclusion criteria as defined by the nine clinical trials. The number of oocytes retrieved was 9.10 ± 5.34 in the patients that met the inclusion criteria (n = 97) versus 6.90 ± 5.23 (P = 0.00122) in those that did not (n = 168). Most 'real life' patients who come for treatment at a tertiary IVF centre do not meet the consensus of inclusion and exclusion criteria used for major clinical trials.

Introduction

In order for an investigational new drug to be approved by the FDA it must first undergo rigorous trials to determine its efficacy and safety. It first enters a preclinical phase where it is tested on animals, followed by three phases of clinical testing on humans. Phase 1 testing is to determine the safety of a drug. It is generally done on a small group (20–100 humans). Phase 2 testing examines the effectiveness, dosing and safety of a drug and is conducted among larger

* Corresponding author.
E-mail address: skol@rambam.health.gov.il (S Kol).

http://dx.doi.org/10.1016/j.rbmo.2016.11.010
1472-6483/

Please cite this article in press as: Eliyakim Hershkop,Linoy Segal,Ofer Fainaru, Shahar Kol, 'Model' versus 'everyday' patients: can randomized controlled trial data really be applied to the clinic?, Reproductive BioMedicine Online (2016), doi: 10.1016/j.rbmo.2016.11.010

groups (100–300 participants). Phase 3 trials are generally large-scale, multiple-site randomized control trials (RCTs). These usually involve thousands of participants (Lipsky and Sharp, 2001).

Additional (Phase 4) trials are often conducted. These examine the real-world effectiveness of a drug in a real-world setting. These trials complement the efficacy data that emanates from pre-marketing RCTs. These real world data may indicate a need for further evaluation via the RCT route or even result in regulatory action (Suvarna, 2010).

The RCT is an essential tool in the development of evidence-based medicine (Sackett et al., 2000). The RCT uses strict inclusion and exclusion criteria as well as a control group, randomization and blinding. These help determine whether given results are indeed an outcome of the specific intervention being examined (Stanley, 2007). This gives the RCT a high level of internal validity and, as such, it is considered the gold standard in clinical trials (Saturni et al., 2014).

Among the essential features of the RCT are the inclusion and exclusion criteria. These criteria limit the subject population in such a way as to strengthen the internal validity of the trial. This results in a more clearly defined group of patients to whom the study results are relevant. Alternatively, the more stringent the criteria, the less able are we to extrapolate the results to patients who do not conform to the predetermined criteria (Saturni et al., 2014). This weakens the external validity (Steckler and McLeroy, 2008).

Van Spall et al. (2007) examined the nature and extent of exclusion criteria among RCTs published in major journals. It was shown that, often, generalizability of results is impaired and that exclusion from trials is unjustified. Reasons for exclusion from trials included common medical conditions, age and receiving commonly prescribed medications and conditions unique to women. The authors concluded that only 47.2% of exclusions in these trials were graded as strongly justified in the context of the specific RCT; most exclusions were deemed poorly justified.

This means that most pharmaceutical study designs result in the exclusion of a large portion of the population, a group that eventually will be treated with the drugs in question. Instead, the studies tend to focus on the 'super-model patient'. This is typically healthy white men aged between 18 and 30 years with no history of comorbidity and and not taking any other medications. As a result, it is often difficult to extrapolate to the 'real life' patient whom the doctor encounters in the clinic (Rothwell, 2010).

This situation undermines the applicability of the RCT and evidence-based medicine paradigm. The obvious caveat of these practices is *a priori* biased evidence-based medicine and the problematic extrapolation from one population to another. These practices have already been challenged for gender (Holdcroft, 2007) and race (Taylor and Wright, 2005).

A pharmaceutical-driven culture of testing drugs only on males and then extrapolating to women (despite significant biological and physiological differences) has been documented (Holdcroft, 2007). Between 1995 and 2005, of 10 drugs withdrawn from the US market, eight were withdrawn as a result of increased danger to women (Simon, 2005).

Similarly, unjustified extrapolation from a homogeneous population to minorities is a common practice. The importance of targeted studies for minorities have been shown in studies such as the African–American Heart Failure Trial (A-HeFT), the African–American Study of Kidney Disease and Hypertension (AASK) and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). (Taylor and Wright, 2005).

We reasoned that the same practice of extrapolating data between groups is pertinent also in the field of infertility and IVF. We hypothesize that drugs used for ovarian stimulation were tested on a narrow population. We also hypothesize that a significant proportion of the patients who come to our clinic do not meet the criteria used for inclusion in previously published pivotal RCTs.

Materials and methods

We surveyed all 305 consecutive patients treated in a tertiary centre IVF Unit during the year 2015. Adequate data were available for 265 out of 305 patients. The demographic and clinical parameters of the 265 patients were compared with the inclusion and exclusion criteria used in nine major clinical trials: Puregon versus Metrodin HP (Out et al., 1995), Gonal-F versus Metrodin HP (Bergh et al., 1997), The Ganirelix Dose-Finding Study Group (1998), Puregon dose (Out et al., 1999), The French Multicentre Trialists Gonal-F versus Metrodin HP (Frydman et al., 2000), The Feronia and Apis study group Gonal-F versus Metrodin HP (Schats et al., 2000), highly purified HMG versus recombinant FSH (Andersen et al., 2006), ENGAGE study (Devroey et al, 2009) and Bemfola versus Gonal-F (Rettenbacher et al., 2015).

Most (90%) of our patients were stimulated using a GnRH antagonist-based protocol. Ovarian stimulation was started on day 2 of the cycle or after scheduling with 4 mg oestradiol. Gonadotrophins were injected daily (type and dose were decided on an individual basis). Daily injections of GnRH antagonist, either 0.25 mg Orgalutran (MSD, Kenilworth NJ, USA) or 0.25 mg Cetrotide (Merck, Darmstadt, Germany), were added when the leading follicle reached 14 mm in diameter until ovulation trigger.

A long GnRH agonist protocol was used in 10% of the patients. Daily injections of 0.1 mg Triptorelin (Ferring, Saint-Prex, Switzerland) were started on day 21. Ovarian stimulation was started approximately 14 days later after ascertaining pituitary down-regulation.

The average total dose of gonadotrophin per cycle in our patients was 1952 ± 776 units. In the eight major gonadotrophin trials, the average doses (IU) were: 2138 versus 2385 (Out et al., 1995); 1643 ± 383 versus 2393 ± 1005 (Bergh et al., 1997); 1114 versus 1931 (Out et al., 1999); 2070 ± 765 versus 3053 ± 1020 (Frydman et al., 2000); 1695 ± 375 versus 1823 ± 383 (Schats et al., 2000); 2508 ± 729 versus 2385 ± 622 (Andersen et al., 2006); 150 µg corifollitropin + 400 IU rFSH versus 1800 IU rFSH (Devroey et al, 2009) and 1556 ± 293 versus 1569 ± 259 (Rettenbacher et al., 2015).

Six inclusion criteria of these different studies were compared and a consensus created. The inclusion criteria were as follows: age, body mass index (BMI), regular cycles, no polycystic ovary syndrome (PCOS), basal FSH levels and basal antral follicle count (AFC). The inclusion consensus for these factors were age 18–38 years, BMI of 18–30, no PCOS, regular cycles, basal FSH levels less than 12 IU/l and basal AFC 20 or over.

For the two categories of patients (those who met the inclusion criteria and those who did not) the following were compared: the number of oocytes retrieved, fertilization rate, embryos created, embryos transferred, implantation rate, gestational sacs and viable embryos. Microsoft Excel spreadsheets were used for data analysis. OriginPro 8.5 (OriginLab USA) was used for statistical analysis. Means and standard deviations were calculated. Continuous variable distributions were compared using the two sided student's t-test. $P < 0.05$ was considered statistically significant.

The study was approved by the Rambam Health Care Campus IRB on 12 July 2015 (study protocol number 0255-15-RMB).

Please cite this article in press as: Eliyakim Hershkop,Linoy Segal,Ofer Fainaru, Shahar Kol, 'Model' versus 'everyday' patients: can randomized controlled trial data really be applied to the clinic?, Reproductive BioMedicine Online (2016), doi: 10.1016/j.rbmo.2016.11.010