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

Results

Only 97 of 265 (37%) patients met the consensus inclusion criteria as defined by the nine clinical trials. A total of 168 out of 265 (63%) did not meet the criteria (**Table 1**). The status of the remaining 40 patients (of the 305) could not be determined owing to incomplete patient records.

Treatment results of patients who did not meet the inclusion criteria were compared with those who met the criteria. The number of oocytes retrieved per patient was 9.10 ± 5.34 in the 97 patients who met the inclusion criteria compared with 6.90 ± 5.23 in the 168 who did not (P = 0.00122). The number of normal fertilizations was 5.01 ± 0.37 in the patients who met the inclusion criteria compared with 3.93 ± 0.29 in the patients who did not (P = 0.02476). No significant differences were detected between the groups in the number embryos created, embryos transferred, implantation rate, gestational sacs and viable embryos (**Table 2**).

Discussion

Drug trial RCTs have intrinsic limitations. As such, post-marketing surveillance studies are needed to determine the real-world effect of the drugs. These studies are often observational and may be retrospective. Although observational studies have limited validity (depending on the appropriateness of design and control for bias), they have the benefits of maximized generalizability and that they are conducted in a real-world environment. Post-marketing drug-utilization studies can be used to examine the relationship between recommended and actual clinical practice. These real-world data can result in important new drug studies, evaluations and recommendations (Suvarna, 2010).

We conducted a small exploratory study consisting of all the patients treated at a tertiary centre IVF Unit in a given year. We believe that our study can provide a platform for further studies on the subject.

A major limitation of our study is that it is a retrospective study. As such, we are comparing retrospective data with prospective RCTs. This limits the statistical value of our study when comparing the inclusion and exclusion criteria of the RCTs to our 'real life' patients. Nevertheless, this was only a secondary outcome of this study and, although limited, it gives us a gross approximation that enables us to compare the two different categories of data.

We chose nine studies that we believe represent the approach to inclusion and exclusion criteria used by major studies supported by the pharmaceutical industry in the field of IVF. We chose six different exclusion and inclusion criteria that the nine aforementioned studies had in common. The six criteria are age, BMI, no PCOS, regular menses, basal FSH levels and basal AFC (**Table 3**). These were chosen because they were the most reliably documented on our database. When a difference in cut-off parameters between the studies was identified, we chose parameters that were closest to the median.

Additional inclusion and exclusion criteria were not used in our study; however, these pertain to a significant number of patients whom we treat in our clinic. These are previous IVF cycles, and poor response in previous IVF cycle (**Table 4**). These were omitted owing to insufficient documentation and inadequate definitions.

One limitation of our study is that, although most of the patients met individual inclusion criteria (**Table 1**), relatively few patients met all six inclusion criteria. The more inclusion criteria that are used, the fewer the number of patients that will meet all of them.

We found that the 'inclusion group' had a higher number of oocytes retrieved and a higher number of normal fertilizations. These results seem to indicate that the RCTs are selected for patients with a better prognosis. Although a greater number of oocytes retrieved usually indicates a higher pregnancy rate, we did not have a large enough sample size to determine whether this was indeed the case in our population.

We found that the 'super-model patient', as described above, was applied, with obvious modifications, to the world of pharmaceutical-initiated ovarian stimulation clinical studies. Indeed, the ENGAGE study (Devroey et al, 2009) notes that the validity of the claims of the study are limited to the study population only. Consequently, about two-thirds of our 'every-day' patients were excluded from these studies.

Of note is that some clinical trial papers did not have clear inclusion and exclusion criteria. This is especially true of The Ganirelix Dose-Finding Study Group (1998) study. Additionally, some studies use vague criteria such as 'normal' FSH instead of using actual values. This makes it more difficult for clinicians to apply the study findings to real life treatment decisions (Rothwell, 2005).

Our study and analysis raises two main concerns: strict adherence to RCT exclusion criteria may lead to denial of useful treatment to potential patients; and exclusion criteria notwithstanding, clinicians often prescribe these treatments for patients who do not meet the inclusion criteria. These clinicians often have insufficient data to guide them. Instead, they must use their best judgment or use the 'educated guess' approach. This could be harmful to patients.

For example, daily practice suggests that the recommended GnRH antagonist daily dose (0.25 mg for both available products (Ganirelix and Cetrorelix) may be insufficient for patients over 40 years of age, or for patients with high BMI. No large RCTs have tested this.

The other side of the coin is a challenge to the pharmaceutical industry: if clinicians are indeed prescribing these medications to patients who were excluded from trials, and the drug works properly (i.e. just as well as with the 'model patient' cohort), why not expand the study criteria to include more 'real life' patients?

Because most of our patient population would be excluded from pivotal RCTs, clinicians should be aware of the limitations inherent in the conclusions of these trials. It may also be ethically and

Table 1 – Number of patients that met the consensus inclusion criteria as defined by the nine clinical trials.

Criterion (n = 265)	$Age_{(years)}$ (18–38)	$BMI_{(kg/m^2)}$ (18–30)	No PCOS	$FSH_{(IU/L)}$ (<12)	$AFC_{(Follicles)}$ (≤20)	Number meeting all consensus criteria
Included (mean values)	167 (35.9±7.1)	192 (26.2±6.0)	229	246 (7.1±3.1)	221 (11.7±9.0)	97
Excluded (mean values)	98 (30.5±4.9)	73 (23.1±2.9)	36	19 (6.6±2.0)	44 (10.7±3.3)	168

Data given as mean ± SEM.
AFC, antral follicle count; BMI, body mass index; PCOS, polycystic ovary syndrome.

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

Table 2 – Reproductive outcome parameters: comparison between included and excluded patients.

Number of oocytes retrieved		Number of normal fertilizations		Number of embryos created		Number of embryos transferred		Implantation rate (%)		Gestational sac (%)		Viable embryos (%)	
Inclusion	Exclusion	Inclusion	Exclusion	Inclusion	Exclusion	Inclusion	Exclusion	Inclusion	Exclusion	Inclusion	Exclusion	Inclusion	Exclusion
9.10 ± 5.34	6.90 ± 5.23	5.01 ± 0.37	3.93 ± 0.29	2.91 ± 0.22	2.67 ± 0.19	1.68 ± 0.07	1.57 ± 0.07	16 ± 3	13 ± 2	28 ± 6	23 ± 4	26 ± 6	23 ± 4
**P = 0.00122		*P = 0.02476		P = NS		P = NS		P = NS		P = NS		P = NS	
n = 97	n = 168	n = 89	n = 153	n = 97	n = 168	n = 97	n = 168	n = 95	n = 168	n = 95	n = 168	n = 95	n = 168

Data given as mean ± SEM.
P values by unpaired Student's t-test.

Table 3 – Six consensus criteria derived from the nine studies.

	Puregon versus Metrodin HP (Out et al., 1995)	Gonal-F versus Metrodin HP (Bergh et al., 1997)	Ganirelix (The Ganirelix Dose-Finding Study Group, 1998)	Puregon (Out et al., 1999)	Gonal-F vs Metrodin HP (Frydman et al., 2000)	Gonal-F vs Metrodin HP (Schats et al., 2000)	hMG vs rFSH (Andersen et al., 2006)	Corifollitropin alpha vs rFSH (Devroey et al, 2009)	Bemfola vs Gonal-F (Rettenbacher et al., 2015)	Consensus criteria
Age (years)	18–39	18–38	18–39	18–39	18–38	18–38	21–37	18–36	20–38	18–38
BMI (kg/m^2)	80–130% Normal BMI	≤28	18–29	18–29	≤30	18–28	18–29	18–32	18–30	18–30
Menstrual cycle (days)	24–35	25–35	24–35	24–35	25–35	25–35	21–35	24–35		Regular cycles
Presence of PCOS	No PCOS	No PCOS	No PCOS	No PCOS	No PCOS	No PCOS	NO PCOS	NO PCOS	No PCOS	No PCOS
Basal FSH (IU/L)					Normal	Normal	≤12	<12	<10	<12
Basal AFC (follicles)		<20			<20			≤20	10–25	≤20

Table 4 – Additional exclusion criteria (not in included in analysis).

	Puregon versus Metrodin HP (Out et al., 1995)	Gonal-F versus Metrodin HP (Bergh et al., 1997)	Ganirelix (The Ganirelix Dose-Finding Study Group, 1998)	Puregon (Out et al., 1999)	Gonal-F versus Metrodin HP (Frydman et al., 2000)	Gonal-F versus Metrodin HP (Schats et al., 2000)	hMG versus rFSH (Andersen et al., 2006)	Corifollitropin alpha versus rFSH (Devroey et al., 2009)	Bemfola versus Gonal-F (Rettenbacher et al., 2015)
Previous IVF cycles	≤3 (oocytes collected ≥1 cycle)	>3 unsuccessful			>3 unsuccessful	≤2 unsuccessful	>3 unsuccessful	>3 unsuccessful	>2
Poor response in IVF cycle		<3 oocytes retrieved in previous cycle		<3 oocytes retrieved in previous cycle	<3 oocytes retrieved in previous cycle	Previous failure	<4 oocytes retrieved in previous cycle or limited follicular response	<4 oocytes retrieved in previous cycle or limited follicular response	<5 oocytes retrieved in previous cycle

medico-legally imperative for clinicians to explain these limitations to their patients.

Acknowledgement

The authors would like to thank Sheldon Hershkop MD and Jason Lefkowitz PhD for their assistance in preparing this paper.

ARTICLE INFO

Article history:
Received 11 June 2016
Received in revised form 25 November 2016
Accepted 25 November 2016
Declaration: The author reports no financial or commerical conflicts of interest.

Keywords:
Evidence-based medicine
Exclusion criteria
Inclusion criteria
IVF
Randomized controlled trials

REFERENCES

Andersen, A.N., Devroey, P., Arce, J.C., 2006. Clinical outcome following stimulation with highly purified hMG or recombinant FSH in patients undergoing IVF: a randomized assessor-blind controlled trial. Hum. Reprod. 21, 3217–3227.

Bergh, C., Howles, C.M., Borg, K., Kamberger, L., Josefsson, B., Nilsson, L., Wikland, M., 1997. Recombinant human follicle stimulating hormone (r-hFSH; Gonal-F) versus highly purified urinary FSH (Metrodin HP): results of a randomized comparative study in women undergoing assisted reproductive techniques. Hum. Reprod. 12, 2133–2139.

Devroey, P., Boostanfar, R., Koper, N.P., Mannaerts, B.M., Ijzerman-Boon, P.C., Fauser, B.C., ENGAGE Investigators, 2009. A double-blind, non-inferiority RCT comparing corifollitropin alfa and recombinant FSH during the first seven days of ovarian stimulation using a GnRH antagonist protocol. Hum. Reprod. 24, 3063–3072.

Frydman, R., Howles, C.M., Truong, F., 2000. A double-blind, randomized study to compare recombinant human follicle stimulating hormone (FSH; Gonal-F) with highly purified urinary FSH (Metrodin) HP) in women undergoing assisted reproductive techniques including intracytoplasmic sperm injection. The French Multicentre Trialists. Hum. Reprod. 15, 520–525.

Holdcroft, A., 2007. Gender bias in research: how does it affect evidence based medicine? J. R. Soc. Med. 100, 2–3.

Lipsky, M.S., Sharp, L.K., 2001. From idea to market: the drug approval process. J. Am. Board Fam. Pract. 14, 362–367.

Out, H.J., Mannaerts, B.M., Driessen, S.G., Bennink, H.J., 1995. A prospective, randomized, assessor-blind, multicentre study comparing recombinant and urinary follicle stimulating hormone (Puregon versus Metrodin) in in-vitro fertilization. Hum. Reprod. 10, 2534–2540.

Out, H.J., Lindenberg, S., Mikkelsen, A.L., Eldar-Geva, T., Healy, D.L., Leader, A., Rodriguez-Escudero, F.J., Garcia-Velasco, J.A., Pellicer, A., 1999. A prospective, randomized, double-blind clinical trial to

study the efficacy and efficiency of a fixed dose of recombinant follicle stimulating hormone (Puregon) in women undergoing ovarian stimulation. Hum. Reprod. 14, 622–627.
Rettenbacher, M., Andersen, A.N., Garcia-Velasco, J.A., Sator, M., Barri, P., Lindenberg, S., van der Ven, K., Khalaf, Y., Bentin-Ley, U., Obruca, A., Tews, G., Schenk, M., Strowitzki, T., Narvekar, N., Sator, K., Imthurn, B., 2015. A multi-centre phase 3 study comparing efficacy and safety of Bemfola(®) versus Gonal-f(®) in women undergoing ovarian stimulation for IVF. Reprod. Biomed. Online 30, 504–513.
Rothwell, P.M., 2005. External validity of randomised controlled trials: 'to whom do the results of this trial apply?'. Lancet 365, 82–93.
Rothwell, P.M., 2010. Commentary: external validity of results of randomized trials: disentangling a complex concept. Int. J. Epidemiol. 39, 94–96.
Sackett, D.L., Straus, S.E., Richardson, W.S., Rosenberg, W., Haynes, R.B., 2000. Evidence-Based Medicine: How to Practice and Teach EBM, 2nd ed. Churchill Livingstone, Edinburgh, pp. 173–177.
Saturni, S., Bellini, F., Braido, F., Paggiaro, P., Sanduzzi, A., Scichilone, N., Santus, P.A., Morandi, L., Papi, A., 2014. Randomized Controlled Trials and real life studies. Approaches and methodologies: a clinical point of view. Pulm. Pharmacol. Ther. 27, 129–138.
Schats, R., Sutter, P.D., Bassil, S., Kremer, J.A., Tournaye, H., Donnez, J., 2000. Ovarian stimulation during assisted reproduction treatment: a comparison of recombinant and highly purified urinary human FSH. On behalf of The Feronia and Apis study group. Hum. Reprod. 15, 1691–1697.
Simon, V., 2005. Wanted: women in clinical trials. Science 308, 1517.
Stanley, K., 2007. Design of randomized controlled trials. Circulation 115, 1164–1169.
Steckler, A., McLeroy, K.R., 2008. The importance of external validity. Am. J. Public Health 98, 9–10.
Suvarna, V., 2010. Phase IV of drug development. Perspect. Clin. Res. 1, 57–60.
Taylor, A.L., Wright, J.T., 2005. Should ethnicity serve as the basis for clinical trial design? Importance of race/ethnicity in clinical trials: lessons from the African-American Heart Failure Trial (A-HeFT), the African-American Study of Kidney Disease and Hypertension (AASK), and the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). Circulation 112, 3654–3660.
The Ganirelix Dose-Finding Study Group, 1998. A double-blind, randomized, dose-finding study to assess the efficacy of the gonadotrophin-releasing hormone antagonist ganirelix (Org 37462) to prevent premature luteinizing hormone surges in women undergoing ovarian stimulation with recombinant follicle stimulating hormone (Puregon). Hum. Reprod. 13, 3023–3031.
Van Spall, H.G., Toren, A., Kiss, A., Fowler, R.A., 2007. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. JAMA 297, 1233–1240.

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