VCWI

Vlaamse Commissie voor Wetenschappelijke Integriteit
Flemish Commission for Research Integrity

Second Advice 2022-01

Dossier VUB / Abu Hashim / Mol

22 December 2022

Complaint

The complaint concerns the PhD thesis *Ovulation induction in polycystic ovary syndrome (PCOS): An appraisal of different strategies* by Prof. dr. Hatem Abu Hashim at the VUB in 2013-'14. "The thesis was based on papers previously published by Prof. Hashim." The complaint mentions serious statistical flaws suggesting data fabrication (duplication of tables, improper randomisation, last digit preference).

The content of the complaint is inspired by the Elsevier publication (editorial, 12p) E.M. Bordewijk et al, Data Integrity of 35 randomized controlled trials in women's health, *Eur J Obstet Gynecol* 249, 72 (2020). In it, Hashim's papers are analyzed and he is accused of fraud. Prof. Ben Mol (Department of Obstetrics & Gynaecology, Monash University, Australia) is the last author of this editorial and he filed the complaint at VUB.

Procedure at the CWI VUB (commission for research integrity)

17 September 2021:	CWI VUB is notified of allegations in a report
16 November 2021:	CWI VUB informs Hashim of allegations
(Meantime)	VUB statisticians Cools and Barbé write an integrity assessment report
14-21 February 2022:	Mails between CWI VUB & Hashim (he is unable to participate in the hearing)
21 April 2022:	Draft CWI Advice was sent (7 days to respond)
15 June 2022:	CWI Advice dated and sent

Factual context

Hatem Abu Hashim is said to have conducted nine Randomized Controlled Trials at Mansoura University, wrote A1 publications about them and received a PhD from the VUB on the basis of those articles (articles = thesis chapters). He has never had an appointment at the VUB.

An Australian-Dutch group (Esmée Bordewijk, Ben Mol) claims to have discovered gross irregularities in these articles (thesis chapters) and analyzed them thoroughly in the aforementioned editorial (March 2020), which forms the basis for the complaint.

In his defense, Hashim wrote "Unequivocal Evidence" (10 page PDF with reaction and 20 PDFs as appendices, a.o. about fabrication detection), in which he criticizes the use of Monte Carlo simulations and the "Carlisle method" and accuses the editorial itself of falsification and other infringements.

Further correspondence about this has been published in the Elsevier journal, including a reaction by Abu Hashim (July 2020) and a counter reaction by Bordewijk, Mol and colleagues (August 2020).

Overview of PhD thesis chapters and sections

Section	Title (= article title)	Reference
1	Introduction	
2.1	Minimal stimulation or clomiphene citrate as first line therapy in women with polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, Bazeed M, Abd Elaal I. *Gynecological Endocrinology*, 2012; 28(2):87-90.	[a]
2.2	Intrauterine insemination versus timed intercourse with clomiphene citrate in polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, Ombar O, Abd Elaal I. *Acta Obstetricia et Gynecologica Scandinavica*, 2011; 90(4):344-50.	[b]
3.1	Three decades after Gjönnaess's laparoscopic ovarian drilling for treatment of PCOS; what do we know? An evidence-based approach. Abu Hashim H, Al-Inany H, De Vos M, Tournaye H. *Archives of Gynecology* and Obstetrics, 2013;288(2):409-22.	Literature study - no study with own data
3.2	Impact of obesity on reproductive outcomes after ovarian ablative therapy in PCOS: A collaborative meta-analysis of individual patient data. Baghdadi LR, Abu Hashim H, Amer SA, Palomba S, Falbo A, Al-Ojaimi E, Ott J, Zhu W, Fernandez H, Nasr A, Ramzy AM, Clark J, Doi SA. *Reproductive BioMedicine Online*, 2012;25(3):227-41.	Systematic review / meta-analysis
3.3	Does laparoscopic ovarian diathermy change clomiphene-resistant PCOS into clomiphene-sensitive? Abu Hashim H, El-Shafei M, Badawy A, Wafa A, Zaglol H. *Arch Gynecol Obstet.* 2011 Aug; 284(2):503-7.	[c]
3.4	Laparoscopic ovarian diathermy after clomiphene failure in polycystic ovary syndrome. Is it worthwhile? A randomized controlled trial. Abu Hashim H, Foda O, Ghayaty E, Elawa A. *Archives of Gynecology and Obstetrics*, 2011; 284(5):1303-9.	[d]
4.1	Combined metformin and clomiphene citrate versus laparoscopic ovarian diathermy for ovulation induction in clomiphene-resistant women with polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, El Lakany N, Sherief L. *Journal of Obstetrics and Gynaecology Research*, 2011;37(3):169-77.	[e]
4.2	Combined metformin and clomiphene citrate versus highly purified FSH for ovulation induction in clomiphene-resistant PCOS women: a randomised controlled trial. Abu Hashim H, Wafa A, El Rakhawy M. *Gynecological Endocrinology*, 2011;27(3):190-6.	[f]
5.1	Letrozole versus laparoscopic ovarian diathermy for ovulation induction in clomiphene resistant women with polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, Mashaly AM, Badawy A. *Archives of Gynecology and Obstetrics*, 2010; 282(5):567-71.	[g]
5.2	Letrozole versus combined metformin and clomiphene citrate for ovulation induction in clomiphene-resistant women with polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, Shokeir T, Badawy A. *Fertility and Sterility*, 2010; 94(4):1405-9.	[h] Already retracted
6	N-acetyl cysteine plus clomiphene citrate versus metformin and clomiphene citrate in treatment of clomiphene-resistant polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, Anwar K, El-Fatah RA. *Journal Of Women's Health* (Larchmt), 2010;19 (11):2043-8	[i]
7-9	Discussion — Summary — Addenda	

Conclusions by CWI VUB

“The VUB CWI concludes that the actions of Prof. Hashim with regard to his thesis constitute a serious violation of scientific integrity, as determined by the Regulations of the VUB CWI (Art. 4, §8), based on the overwhelming evidence of fabrication of statistical outcomes detected by Professors Barbé and Cools, and on the clear lack of statistical proficiency. The thesis submitted by Professor Hashim does not merit the award of the title of Doctor of Medical Sciences. [..] The VUB CWI advises the Rector to continue this investigation via the appropriate internal disciplinary bodies.”

The report by CWI VUB relies on an expert analysis by prof. Kurt Barbé (mathematician) and Wilfried Cools (statistician), both connected to the Medical Informatics Team of the VUB Faculty of Health Sciences. In their report, they identify four different problems or integrity infringements:

1. Defective statistics: all tests are incorrect (although only simple two sample t-tests were reported), statistical characteristics are inconsistent (although the tests do not require more elaborate knowledge than bachelor-level statistics).
2. Non-randomized groups or undeclared reuse of subjects: very similar baseline characteristics, each differing by at most one digit.
3. Fabrication of statistical values: inconsistencies between data properties, p-values, confidence intervals, test characteristics, showing that they are not (correctly) calculated from the data. Last digit p-values in [4,9] only.
4. Fabrication of data: the assumption that an original dataset exists is difficult to maintain.

The report judges 1 and 3 proven. There is strong evidence for 2. Fabrication (4) is only provable with the dataset.

Procedure at VCWI

8 July 2022:	Request for a second advice sent by Hatem Abu Hashim (accused).
1 September 2022:	First discussion and decision to handle the case.
13 October 2022:	Hearing with Ben Mol, hearing with Hatem Abu Hashim, hearing with John Pearson and Paul Geerlings (CWI VUB), second deliberation.
24 November 2022:	Discussion of the draft second advice, discussion on the qualification.
22 December 2022:	Adoption of the second advice.
23 December 2022:	Second advice is sent.

Findings and considerations of the VCWI

In their investigation regarding the complaint of data falsification and/or data fabrication by Abu Hashim, the VCWI found the following.

No rebuttal to the allegation

The VCWI notices that strikingly, Hatem Abu Hashim does not defend his work against any of the allegations. When faced with such a serious allegation, one could expect that the accused would (1) address the clear statistical errors (admit the mistakes as sloppy science, or show why they're not erroneous); and (2) try everything possible to prove that the trials have been executed, by showing any sort of evidence.

In his lengthy reply, Abu Hashim does not address the allegations at all. To the contrary, his defence consists mainly of accusing those bringing forward the complaint of misconduct and questioning their work and methods.

This seems to be a pattern, as Abu Hashim's work was flagged to be problematic, first by Bordewijk et al. (2020), Nick Brown's blogpost (2021) and the CWI VUB report (2022). After each of these reports, the VCWI has found no meaningful rebuttal by Abu Hashim refuting the evidence or proving that the studies have actually taken place. Also in the hearing before the VCWI, Abu Hashim presented no evidence that would clear him from the allegations.

Statistical nonsense

The VCWI found, following evidence presented by Bordewijk et al. (2020), Nick Brown's 28 October 2021 blogpost and the CWI VUB expert report by Kurt Barbé and Wilfried Cools, that there are many gross statistical errors in the papers. For example, as Nick Brown noted in his blogpost, the paper [e], being section 4.1 of the PhD thesis, contains this table, with the unpaired group difference for fasting glucose between 92.5 ± 1.8 (n=95) and 87.1 ± 1.1 (n=43) impossibly being non-significant. In fact, one can calculate that the pooled standard deviation would be around 1.63, the *t* characteristic would be around 18 and the actual p value would be $p < .00001$.

Table 3 Baseline characteristics of ovulatory and anovulatory patients in each group

	Group A (combined metformin–CC) ($n = 138$)		
	Ovulation ($n = 95$)	No ovulation ($n = 43$)	*P*-value
Age (years)	26.6 ± 1.7	27.8 ± 1.6	NS
Duration of infertility (years)	3.7 ± 1.1	4.1 ± 1.4	NS
BMI (kg/m^2)	24.8 ± 1.7	31.8 ± 2.2	0.003
Waist-to-hip ratio	0.8 ± 0.01	0.8 ± 0.02	NS
Testosterone (ng/mL)	0.72 ± 0.01	1.6 ± 0.2	0.03
SHBG (nmol/L)	34.2 ± 2.3	20.4 ± 2.2	0.02
FAI (%)	5.6 ± 2.1	12.6 ± 3.5	0.01
LH (mIU/mL)	12.8 ± 2.8	11.5 ± 3.1	NS
FSH (mIU/mL)	5.4 ± 0.3	5.7 ± 0.6	NS
LH/FSH ratio	2.7 ± 1.2	2.5 ± 1.1	NS
Fasting glucose (mg/dL)	92.5 ± 1.8	87.1 ± 1.1	NS
Fasting insulin (μU/mL)	16.7 ± 2.2	12.1 ± 1.5	0.001
Fasting glucose/insulin ratio	5.8 ± 2.1	8.7 ± 2.6	0.001
Ovarian volume (mL)	11.5 ± 1.4	11.4 ± 1.3	NS

In paper [c], being section 3.3 in the thesis, the exact same values are presented (albeit for different groups of patients in another study, remarkably with precisely the same values).

	LOD non-responders ($n = 84$)		
	Ovulation ($n = 30$)	No ovulation ($n = 54$)	P value
Age (years)	26.4 ± 1.7	27.8 ± 1.6	0.96
Duration of infertility(years)	3.7 ± 1.1	5.2 ± 1.6	0.87
BMI (kg/m^2)	24.8 ± 1.7	32.3 ± 2.2	<0.001[a]
Waist-to-hip ratio	0.8 ± 0.01	0.8 ± 0.02	0.23
Ferriman–Gallwey score	9.1 ± 1.1	12.7 ± 1.4	0.02[a]
Testosterone (ng/mL)	0.73 ± 0.01	1.5 ± 0.2	0.03[a]
SHBG (nmol/L)	34.2 ± 2.3	20.4 ± 2.2	<0.001[a]
FAI (%)	5.6 ± 2.1	12.6 ± 3.5	<0.001[a]
LH (mIU/mL)	12.8 ± 2.8	11.5 ± 3.1	0.85
FSH (mIU/mL)	5.4 ± 0.3	5.7 ± 0.6	0.16
LH/FSH ratio	2.7 ± 1.2	2.5 ± 1.1	0.63
Fasting glucose (mg/dL)	92.5 ± 1.8	87.1 ± 1.1	0.96
Fasting insulin (μU/mL)	12.1 ± 1.5	16.7 ± 2.2	0.003[a]
Fasting glucose/insulin ratio	8.7 ± 2.6	5.8 ± 2.1	0.002[a]
Ovarian volume (mL)	11.5 ± 1.4	11.4 ± 1.3	0.27

A *p*-value of 0.96 is clearly incorrect for such (relative to their SD's) large mean differences.

The same remarkability holds for the Testosterone lines in the same tables. Looking at the first table from paper [e]: if indeed the 95 ovulating women would *all* have testosterone levels of 0.720 ng/mL (± 0.01), the comparison with the non-ovulating group (testosterone = 1.6 ng/mL ± 0.20) could not possibly yield a *p*-value of 0.03. One can calculate that the correct *p*-value would be $4{,}66\times10^{-118}$ (-118 is not a typo).

This is not an isolated case, quite to the contrary. As a sample, we have recalculated 23 *p*-values across the chapters — at least two in each of the nine papers (PhD thesis sections) containing tables. The result is that none of those were correct: none of the reported *p*-values were consistent with the data summary values. Many of our checks yielded that, based on the summary values, *p*-values of the order of magnitude of 10^{-35} and 10^{-60} were no exceptions. These are very unlikely to come from clinical practice, especially when comparing baseline characteristics between two groups emerging from a *random* split.

In all of the thesis chapters except 5.2, data summary tables are like those above: they contain only sample sizes, means, standard deviations and *p*-values. Only in paper [h], included as chapter 5.2 in the thesis, test characteristics are reported. However, none of the *t* statistics and *p*-values are consistent. To give an example, in Table 2 (see next page) of the article [h], section 5.2 in the PhD thesis, the first line exhibits a difference between 4.4 ± 0.4 and 6.8 ± 0.3 in large patients groups of n=123 and n=127. It can be calculated that the *t* statistic must be around 53 and not 4.3. The reported *p*-value of 0.042 is consistent with neither ($t = 4.3$ would yield $p = 0.000025$, $t = 53$ would give $p < 0.00001$).

TABLE 2

Outcome in letrozole and combined metformin–CC groups.

	Group A (letrozole group) (n = 123)	Group B (combined metformin–CC group) (n = 127)	*t*	χ^2	*P* value
Total no. of follicles	4.4 ± 0.4	6.8 ± 0.3	4.3		.042[a]
No. of follicles >14 mm	2.1 ± 0.3	3.7 ± 0.5	6.13		.008[a]
No. of follicles >18 mm	2.3 ± 0.1	3.1 ± 0.8	5.03		.003[a]
Pretreatment endometrial thickness	5.5 ± 0.5	5.3 ± 0.6	1.31		.22
Endometrial thickness at hCG (mm)	9.5 ± 0.2	9.1 ± 0.1	1.44		.53
Serum E_2 (pg/mL)	258 ± 62.2	386 ± 88.3	4.12		.022[a]
Serum P (ng/mL)	7.3 ± 0.9	11.4 ± 1.2	6.33		.024[a]
Duration of stimulation (d)	12.2 ± 1.3	8.1 ± 2.8	4.91		.036[a]
Ovulation/cycle	185/285 (64.9%)	207/297 (69.6%)		1.63	.82
Pregnancy/cycle	42/285 (14.7%)	43/297 (14.4%)		1.32	.53
Miscarriage/patient	4 (10.2%)	4 (9.5%)		1.73	.43

[a] Statistically significant difference: $P < .05$.

Abu Hashim. Letrozole or CC-metformin for CC resistance. Fertil Steril 2010.

In the same paper, table 1 shows *t* characteristics all between 2.66 and 3.90. Knowing that a *t* distribution with 123+127-2=248 degrees of freedom closely resembles a normal distribution, that a characteristic of 1.96 matches a *p*-value of 0.05 and all *t* characteristics higher than 1.96 should yield *p*-values less than 0.05, it is clear that the table below doesn't have a single correct row with *t* values.

TABLE 1

Patients' characteristics.

	Group A (letrozole group) (n = 123)	Group B (combined metformin–CC group) (n = 127)	*t*	χ^2	*P* value
No. of cycles	285	297			
Age (y)	28.3 ± 2.7	26.2 ± 2.2	2.66		.1
Parity					
Nulliparous	100 (81.3%)	106 (83.5%)		0.28	.83
Multiparous	23 (18.7%)	21 (16.5%)		0.87	.48
Clinical presentation					
Oligo/anovulation	90 (73.1%)	94 (74%)		0.03	.86
Hyperandrogenism	51 (41.5%)	60 (47.2%)		0.99	.32
Polycystic ovaries	88 (71.5%)	88 (69.2%)		0.26	.60
Height (cm)	166.3 ± 5.2	162.1 ± 5.1	2.95		.08
Weight (kg)	85.3 ± 6.4	88.1 ± 4.2	2.88		.09
BMI (kg/m^2)	29.1 ± 3.2	30.1 ± 2.3	2.05		.83
FSH (IU/mL)	4.6 ± 2.1	5.1 ± 2.2	3.88		.06
LH (IU/mL)	11.4 ± 2.8	11.8 ± 2.6	3.90		.06

Note: None of the differences were statistically significant ($P > .05$).

Abu Hashim. Letrozole or CC-metformin for CC resistance. Fertil Steril 2010.

Rather than "statistical errors", these gross inconsistencies and unexplainable values, can better be termed as "statistical nonsense", as the numbers appear to be quite random.

These findings confirm the conclusions of the Barbé-Cools report (which informed the VUB CWI advice) that "the reported characteristics of the sample mean, standard deviation, *t*-statistics/chi-squared statistics and *p*-values are inconsistent" and that "none of the statistical analyses are correct".

Abu Hashim states in all of his thesis sections (being published papers) that "Data obtained were statistically analyzed using Statistical Package for Social Sciences (SPSS, Chicago, USA) software version 15.0 for Windows." That is clearly incorrect, as SPSS computes test characteristics and *p*-values automatically and correct. Only two other explanations are reasonable tenable: (1) that the statistical outcomes from the software were changed before submission — this is falsification; or (2) that the statistical outcomes never originated from SPSS software and not even from real medical data, but were made up by the author(s) — this is fabrication.

Unlikely high number of duplicate (or very similar) values

As noted before, some tables contain very similar values, like these (screenshots below are the crops from the red rectangles in the previous section):

Testosterone (ng/mL)	0.72 ± 0.01	1.6 ± 0.2	0.03
Fasting glucose (mg/dL)	92.5 ± 1.8	87.1 ± 1.1	NS
Testosterone (ng/mL)	0.73 ± 0.01	1.5 ± 0.2	0.03[a]
Fasting glucose (mg/dL)	92.5 ± 1.8	87.1 ± 1.1	0.96

These 8 numbers are pairwise "similar" (6 identical, 2 differing one digit). Following Bordewijk et al. (2020), a 'similarity' may be defined as 'means, standard deviation, percentage, *t*-value, *p*-value or confidence interval having the same value in a pair of studies, or +1 or -1 for any digit'. Confirming their work, there is indeed an extremely unlikely high number of similarities between means, standard deviations and, where available, percentages, reported t-values, reported p-values and confidence intervals in the reports of a number of studies that are described as independent randomized control studies in thesis chapter sections 2.1, 2.2, 3.4, 4.1, 4.2, 5.1, 5.2, and 6.

When pairwise comparing all reported values related to *baseline characteristics*, in for example the two "randomized clinical trials" in section 2.1 and 2.2 of the PhD Thesis, 49 of these value pairs are in fact the same number, or differing in one digit (moreover, this differing digit being altered +1 or -1 only). Among these 49 similarities are 10 exactly the same *p*-values. When comparing values related to *outcome variables*, there are 33 similarities. Although value pairs may sometimes coincidentally be the same or very similar, the observed similarities here are extremely prevalent, making them highly suspicious (too prevalent to be credible).

This is not only true for the comparison of the two studies in sections 2.1 and 2.2, but for all reported studies throughout the PhD Thesis. The following tables show the number of similarities among baseline and among outcome variables across included trials.

	Baseline characteristics							*Outcome variables*						
section	2.1	2.2	3.4	4.1	4.2	5.1	5.2	2.1	2.2	3.4	4.1	4.2	5.1	5.2
2.2	49							33						
3.4	29	25						4	2					
4.1	19	21	23					5	2	7				
4.2	28	26	26	36				7	5	3	2			
5.1	21	20	27	41	33			3	4	7	7	3		
5.2	4	1	4	3	2	5		8	11	5	5	7	5	
6	31	25	19	24	29	25	12	6	7	5	3	10	3	19

The number of similarities is extremely high between all studies (with the possible exception of the study reported in section 5.2 when considering baseline characteristics).

The similarities are too high to be credible. At the very least, these unexplainable similarities among result values strongly suggest that the studies were not conducted as described in the papers (i.e., as independent randomized control trial studies). Such striking similarities are not naturally to be expected when studying data from real patients.

Other findings

In their search for clues shedding light on which scenario was most probable, or even provable, the VCWI found the following:

- **No data.** The original data of all of these trials are unavailable. Abu Hashim said in the hearing that the patient data were present in paper form, kept in Mansoura University for 5 years after the publication, in line with the University's data retention policy, with no back-up files. Although the commission considers this position plausible in principle, the fact that more than one thousand infertile women participated in these trials, should have left some proof or trail (for example in the "private practice" which is mentioned in all of the 9 papers alongside "Outpatient Clinic in Mansoura University Hospitals, Egypt", from which patients were both recruited). The commission noticed that Abu Hashim did not make any effort to provide any form of evidence that the studies have been actually conducted.

- **Much recycled.** All Abu Hashim's reports of randomized controlled trials included in the PhD thesis are very similar. They are relatively short and they follow the same structure. The text of the materials and methods section seems to be copied and the acknowledgements are almost always the same.

- **Simple but very concise statistics** Abu Hashim's reports of randomized controlled trials included in the PhD thesis almost all contain a table comparing the baseline characteristics of two groups of patients, and one or more tables comparing the outcomes across those groups. Abu Hashim and co-authors write in their papers "Means were compared using the unpaired Student’s t-test". The statistics are very simple. Nonetheless, the statistical reporting is very concise: values don’t show many significant digits (1 to 2, seldomly 3 digits), *t* statistics are almost never shown. The *p*-values are sometimes hidden as “NS” (some tables include only NS’s with no *p*-value).

- **Suspicious values** In their letter of response (2) of 29 August 2020, Bordewijk et al. remarked that the 40 reported pre-treatment endometrial thickness means across 9 different studies, only the digits 4, 5, 6 and 7 occurred as the last digit for measurement values — while the other numbers (0, 1, 2, 3, 8, 9) never occurred as last digit. Indeed, across the different chapters of the PhD thesis, 6 of the 8 pairwise pre-treatment endometrial thickness means comparisons are 5.4 — 5.6. Given the reported sample sizes (varying between 55 and 144) and standard deviations (that vary between 0.4 and 0.6), it is highly unlikely to find these values if all studies described in the PhD are truly independent randomized control trials.

- **Impossible values** When looking at the values presented in the tables, many values seem too strange to come from real patient data. For example, in paper [e], included as section 4.1 in the PhD thesis, table 3 shows a waist-to-hip ratio in four groups of women, all with mean 0.8 and with standard deviations of 0.01 (twice), 0.02 and 0.03. Waist-to-hip ratio is a dimensionless fraction, calculated by dividing the circumference of waist, measured at the smallest point above the belly button, by the circumference around the hips where it is maximal. The Deutsche Gesellschaft für Sportmedizin und Prävention defines overweight as a waist-to-hip-ratio between 0.80 and 0.85 (less being normal and more being obese), which

gives an idea of the expected variability of this variable. A standard deviation of 0.01 from the mean 0.800 is incredibly impossible, as this would mean that almost all of the 95 participants in the ovulation group (but not exactly all) would have a waist-to-hip ratio of precisely 0.800. This cannot be realistically expected from 95 random women. (In the best case scenario, it would be needed to give some more digits after the digital point of this mean "0.8" when the standard deviation is 0.01. Without these values, it cannot be properly established that the difference is non-significant.)

Table 3 Baseline characteristics of ovulatory and anovulatory patients in each group

	Group A (combined metformin–CC) ($n = 138$)		
	Ovulation ($n = 95$)	No ovulation ($n = 43$)	*P*-value
Age (years)	26.6 ± 1.7	27.8 ± 1.6	NS
Duration of infertility (years)	3.7 ± 1.1	4.1 ± 1.4	NS
BMI (kg/m^2)	24.8 ± 1.7	31.8 ± 2.2	0.003
Waist-to-hip ratio	0.8 ± 0.01	0.8 ± 0.02	NS
Testosterone (ng/mL)	0.72 ± 0.01	1.6 ± 0.2	0.03
SHBG (nmol/L)	34.2 ± 2.3	20.4 ± 2.2	0.02
FAI (%)	5.6 ± 2.1	12.6 ± 3.5	0.01
LH (mIU/mL)	12.8 ± 2.8	11.5 ± 3.1	NS
FSH (mIU/mL)	5.4 ± 0.3	5.7 ± 0.6	NS
LH/FSH ratio	2.7 ± 1.2	2.5 ± 1.1	NS
Fasting glucose (mg/dL)	92.5 ± 1.8	87.1 ± 1.1	NS
Fasting insulin (μU/mL)	16.7 ± 2.2	12.1 ± 1.5	0.001
Fasting glucose/insulin ratio	5.8 ± 2.1	8.7 ± 2.6	0.001
Ovarian volume (mL)	11.5 ± 1.4	11.4 ± 1.3	NS

Is fabrication proven?

The Barbé-Cools expert report of CWI VUB suspected fabrication but did not consider fabrication formally proven, as serious falsification could not be excluded: "[T]to understand at what level the manipulation was conducted, the authors should demonstrate how the results were obtained starting from the study design to the reported statistical outcomes." Abu Hashim did not explain this to the CWI VUB and stopped responding following their draft report. In the hearing for the VCWI, he did not explain either. In fact he hardly referred to his own research but mainly criticized his critics (with arguments that did not make much sense).

The findings give credibility to the scenario that the articles were produced without any data originating from clinical practice and were fabricated completely (at best that only a very limited amount of data were gathered and the remaining invented data were modeled accordingly). All of these findings above make the scenario credible that Abu Hashim had learned to write medical papers by reading others, that he made up all reported values and that he wrote more papers by adapting previous papers, copying results between articles and applying small alterations (+1 or -1 in some digits). This would be a very plausible explanation for the statistical inconsistencies, duplicated (or slightly altered) values, conciseness pertaining to numbers and definitely the striking similarities between the papers' structure, content and text.

However credible and plausible this scenario may be, this forms no proof. The commission has deliberated and discussed thoroughly whether by any chance all the findings above may be reasonably explained by real trials, combined with very sloppy science and lack of statistical proficiency. This would give Abu Hashim the benefit of the doubt, concluding only there are serious doubts but no proof of fabrication.

However, the commission has reached the conclusion that, given the combination of

- the clear statistical nonsense throughout all nine studies in the PhD; and
- the extremely unlikely similarities in reported values across all nine studies in the PhD;

moreover combined with

- the similarity of all the papers, their structure, content and considerable textual overlap;
- the unavailability of any original data for either of all the studies;
- the unavailability of any other evidence that the studies have really taken place;
- the absence of any effort by Abu Hashim to explain the statistical nonsense or to prove that the studies have actually taken place,

the scenario that all values and results indeed originate from medical data from actual clinical trials is no longer reasonably tenable. The conclusion of complete (or virtually complete) fabrication is the only reasonable explanation for the findings. When considering all elements together, the commission finds the presented evidence convincingly demonstrating fabrication.

Even in the very unlikely event that true data would come up, the conclusion of research misconduct (falsification) still stands. Fabrication is making up results and recording them as if they were real. Falsification is manipulating research materials, equipment or processes or changing, omitting or suppressing data or results without justification. (Definitions from the 2017 ALLEA Code of Conduct for Research Integrity) Both are serious infringements of research integrity, distorting the research record. Without access to the underlying medical data, it is impossible to establish whether the reported values were adaptations from the correct values originating from patient data, or were made up completely or copied from other studies. However, the conclusion (fabrication *or* falsification) stays the same in all cases: a serious breach of research integrity.

Procedural matters

In his request for a second advice, Hatem Abu Hashim mentions different issues in the investigation procedure by CWI VUB, summarized here in *italic*.

- *CWI insisted on its own investigation, even though an investigation was ongoing in Abu Hashim's home university (Mansoura).* It is perfectly legitimate that VUB runs its own investigation, as VUB awarded the PhD and the allegation suggests that the work would not merit the PhD title. CWI regulations state that when different Flemish universities are concerned in an allegation, there is a consultation about which CWI will handle the case. This is not applicable here, as Mansoura University is not a Flemish university.
- *Seven days is not enough to respond to the draft advice.* The procedural terms are a choice of the institution — it's not up to the VCWI to judge on this choice. Nevertheless, the VCWI noticed that Hatem Abu Hashim did not ask for prolongation or delay of this term — in fact he did not respond at all since then.
- *The Barbé-Cools report (integrity assessment) is selective, does not address three issues, does not respond to its "unequivocal evidence" and extends the analysis to an RCT beyond the thesis. Moreover, it uses unreliable methods.* As noted earlier, Hatem Abu Hashim questions the analysis of all those scrutinizing his work. The defence "unequivocal evidence" criticises the

Bordewijk-Mol editorial. The Barbé-Cools report rightly addresses the allegation of fabrication with an analysis of the work, independent from the work by Bordewijk and Mol. So the VCWI sees no need for the CWI VUB or its Barbé-Cools report to respond to "unequivocal evidence" or address the issues raised by Abu Hashim. The statistical errors revealed are so evident that no specific methods are needed.

- *The report calls Bordewijk et al. a meta-analysis instead of an editorial.* In the editorial, different trials are analysed to draw conclusions spanning more than one study. However, the difference between both words is not relevant for the judgment of the allegation. The content is.

Judgment by VCWI and second advice

The VCWI judges there is no other reasonable explanation that can come from the strong evidence presented in this report, acknowledging the work previously done by Bordewijk, Mol et al., Nick Brown and Barbé and Cools, than fabrication. Based on the above findings and considerations, the VCWI judges the allegation of fabrication by Hatem Abu Hashim to be proven. This is the most serious form of research misconduct.

It is a matter of concern for the VCWI that a doctoral title can be awarded on the basis of published but fabricated research, with journal peer review as the only quality control.

The commission is grateful for the in-depth investigations by Ben Mol, Esmée Bordewijk, Nick Brown and others, which have brought this case to the surface. The members of the commission have appreciated their time consuming and unpaid work, which is ultimately beneficial for science.

Advice

In the interest of the scientific literature, clinical practice and patient well-being, as well as to set an example, the VCWI recommends the following, to the VUB in the first place.

1. The evidence of fabrication justifies a withdrawal of the PhD title by VUB.

2. The following nine articles should be immediately retracted. Although fabrication by Hatem Abu Hashim is considered proven, it is not 100 % certain that all nine of these studies are fabricated — a slight possibility remains that one or some were based on actual patient data. However, the statistical nonsense and high likelihood of fabrication justify a retraction of all of the following papers featuring in the VUB PhD thesis.

    a. Minimal stimulation or clomiphene citrate as first line therapy in women with polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, Bazeed M, Abd Elaal I. *Gynecological Endocrinology*, 2012; 28(2):87-90.
    b. Intrauterine insemination versus timed intercourse with clomiphene citrate in polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, Ombar O, Abd Elaal I. *Acta Obstetricia et Gynecologica Scandinavica*, 2011; 90(4):344-50.
    c. Does laparoscopic ovarian diathermy change clomiphene-resistant PCOS into clomiphene-sensitive? Abu Hashim H, El-Shafei M, Badawy A, Wafa A, Zaglol H. *Arch Gynecol Obstet.* 2011 Aug; 284(2):503-7.

d. Laparoscopic ovarian diathermy after clomiphene failure in polycystic ovary syndrome. Is it worthwhile? A randomized controlled trial. Abu Hashim H, Foda O, Ghayaty E, Elawa A. *Archives of Gynecology and Obstetrics*, 2011; 284(5):1303-9.
e. Combined metformin and clomiphene citrate versus laparoscopic ovarian diathermy for ovulation induction in clomiphene-resistant women with polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, El Lakany N, Sherief L. *Journal of Obstetrics and Gynaecology Research*, 2011;37(3):169-77.
f. Combined metformin and clomiphene citrate versus highly purified FSH for ovulation induction in clomiphene-resistant PCOS women: a randomised controlled trial. Abu Hashim H, Wafa A, El Rakhawy M. *Gynecological Endocrinology*, 2011;27(3):190-6.
g. Letrozole versus laparoscopic ovarian diathermy for ovulation induction in clomiphene resistant women with polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, Mashaly AM, Badawy A. *Archives of Gynecology and Obstetrics*, 2010; 282(5):567-71.
h. Letrozole versus combined metformin and clomiphene citrate for ovulation induction in clomiphene-resistant women with polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, Shokeir T, Badawy A. *Fertility and Sterility*, 2010; 94(4):1405-9. (This article was already retracted.)
i. N-acetyl cysteine plus clomiphene citrate versus metformin and clomiphene citrate in treatment of clomiphene-resistant polycystic ovary syndrome: a randomized controlled trial. Abu Hashim H, Anwar K, El-Fatah RA. *Journal Of Women's Health* (Larchmt), 2010;19 (11):2043-8

The primary responsibility to request retraction lies with Mansoura University. We advise the VUB (1) to request Mansoura University to start up the retraction process; and moreover (2) in the case where Mansoura University would not respond, to request retraction of the eight papers directly with the journals. Both universities are at liberty to include the VCWI report for this purpose.

3. Specific attention and work is needed to stop the untrustworthy results in the retracted papers from percolating into meta-analyses, medical guidelines and clinical practice.

4. Coordinated and well-thought public communication about the case and its consequences. You might consider to applaud in this public communication the whistleblowers' constructive work in safeguarding research integrity, with their names mentioned if they agree beforehand. Consider to also inform the Royal College of Obstetricians and Gynaecologists, of which Abu Hashim is a member, before public communication.