Clinica Chimica Acta 468 (2017) 114–119

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

Validation of image cytometry for sperm concentration measurement: Comparison with manual counting of 4010 human semen samples

Dorte L. Egeberg Palme [a,b], Trine Holm Johannsen [a,b], Jørgen Holm Petersen [a,b,c], Niels E. Skakkebæk [a,b], Anders Juul [a,b], Niels Jørgensen [a,b], Kristian Almstrup [a,b,*]

[a] Department of Growth and Reproduction, Rigshospitalet, University of Copenhagen, Denmark
[b] International Center for Research and Research Training in Endocrine Disruption of Male Reproduction and Child Health (EDMaRC), Rigshospitalet, University of Copenhagen, Denmark
[c] Department of Biostatistics, Institute of Public Health, University of Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 6 December 2016
Received in revised form 20 January 2017
Accepted 20 February 2017
Available online 27 February 2017

Keywords:
Sperm concentration
Image cytometry
Manual counting
Reproducibility

ABSTRACT

Sperm concentration is an essential parameter in the diagnostic evaluation of men from infertile couples. It is usually determined by manual counting using a hemocytometer, and is therefore both laborious and subjective. We have earlier shown that a newly developed image cytometry (IC) method may be used to determine sperm concentration. Here we present a validation of the IC method by analysis of 4010 semen samples. There was high agreement between IC and manual counting at sperm concentrations above 3 mill/ml and in samples with concentrations above 12 mill/ml the two methods can be used interchangeable. However, we found substantial differences in samples below 3 mill/ml. We also assessed the accuracy of the two methods by repeated measurements of 248 samples, which revealed that IC measurements seemed more accurate. Moreover, based on ten samples counted by several operators the IC method had a lower coefficient of variation than the manual method (5% vs 10%), indicating a better precision of the IC method.
In conclusion, measurement of sperm concentration by IC can be used at concentrations above 3 mill/ml and seems more accurate and precise than manual counting, making it an attractive option in the daily clinical practice.

1. Introduction

Infertility is an important and widespread health issue, and in the diagnostic work-up of the male partner, analysis of the semen is compulsory. Usually, determination of sperm concentration is assessed by the use of a hemocytometer as described by the World Health Organization [1,2]. Manual procedures are, however, labor intensive, based on counting of relatively few spermatozoa and can be biased by high intra- and inter-observer variations [3,4]. Consequently, international quality programs and training courses are offered by international fertility associations on how to analyze semen samples in a standardized manner, and several efforts have been made to automate the procedure. Both flow cytometric and computer assisted sperm analysis (CASA) systems [5–10] have been used to determine human sperm concentration. However, the application of these automated instruments in the daily clinical practice has been limited, probably because large comparative studies have not provided sufficient evidence of their usefulness. We have, however, previously shown, that image cytometry (IC) can be used to determine the sperm concentration with great ease and precision, at least at concentrations above 4 mill/ml [11], and hence the method could be suited for automatic counting of sperm concentration in the clinic.

Here we validate the previous results by presenting new data from measurements of a total of 4010 semen samples analyzed both with the manual hemocytometer method and by the new IC method. Moreover, the reproducibility and accuracy of the two methods by repeated measurements were also investigated.

2. Material and methods

2.1. Semen samples

A total of 4010 semen samples (729 of these were also included in Egeberg et al. [11]) from 2720 men with varying sperm concentrations were included in this study, which was conducted from November 2011 to October 2014. The samples were delivered by men from infertile couples (N = 1666), from an ongoing study of reproductive health in men from the general Danish population (N = 633 men; one sample each) [12], and from other ongoing studies in the department (N = 421, one sample each). All semen samples were produced by masturbation and ejaculated into clean, wide-mouthed plastic containers. The samples

Abbreviation: IC, image cytometer.

* Corresponding author at: Department of Growth and Reproduction, GR-5064, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark.
E-mail address: kristian@almstrup.net (K. Almstrup).

http://dx.doi.org/10.1016/j.cca.2017.02.014
0009-8981/

were collected in the privacy of a room near the laboratory or at home and delivered to the laboratory within a maximum of one hour after ejaculation. After ejaculation, the samples were allowed to liquefy for at least 30 min at 37 °C. Following thorough mixing, aliquots were drawn for assessment of sperm concentration by the two methods.

2.2. Hemocytometric assessment of sperm concentration

The method for manual assessment of sperm concentration has been described previously [11]. Briefly, semen samples were diluted in a solution of 0.6 M $NaHCO_3$-buffer and 0.37% (v/v) formaldehyde in distilled water and counted in replicates using a Bürker-Türk hemocytometer (Paul Marienfeld GmbH & Co. KG, Lauda-Königshofen, Germany). Replicate counts were compared according to Poisson as described by WHO [2]. Unacceptable differences resulted in preparation of new dilutions and repeating of replicate counts. If no spermatozoa were observed the sample was classified as azoospermic, and after centrifugation the pellet was re-investigated for the presence of spermatozoa.

2.3. Automated assessment of sperm concentration by IC

The automated assessment of sperm concentration by IC has been described in detail elsewhere [11]. Briefly, semen samples were diluted in S100-buffer (ChemoMetec A/S, Allerød, Denmark), and immediately loaded into PI-Cassettes™ (ChemoMetec) for assessment in a NC-3000 ™ image cytometer (ChemoMetec). S100-buffer contains a detergent (<2.5% w/w octylphenol ethoxylate) that disrupts the plasma membrane and renders the nuclei susceptible to staining with PI, without dispersing the protamine-packed sperm nuclei. A predefined script was used to acquire images and analyze data. Replicate counts were compared according to Poisson as described in Section 2.2. The same instrument was used for all measurements and IC measurements were performed at the same time after ejaculation as the manual counting. Positive displacement pipettes were used for both methods.

2.4. Recounting of samples

For all samples the manually counted and the automated assessed concentration were compared according to Poisson. If the difference between the two methods was unacceptable (Poisson >1.96) and if the sample volume allowed for recounting, independent additional counts were performed In total 425 out of 4010 samples (11%) differed more than expected between the two methods. 177 could not be recounted because of insufficient volume leaving 248 recounted samples for analysis. 147 samples were re-counted with both methods, 89 only with IC, and 12 only with the manual method. Detailed description can be found in Egeberg et al. [11].

2.5. Inter-observer variation

Inter-observer variation of manual counting and image cytometry was assessed in autumn 2013 and spring 2014. For these anonymous assessments, semen samples were mixed to obtain different concentrations. Ten samples were assessed (as described above) by four technicians by manual counting and by three technicians using the image cytometer. The involved personnel were blinded for each other's measurements. They were only informed about the dilution factor of each sample. The technicians prepared their own dilutions of the samples, and assessments were done in replicate as described Sections 2.2 and 2.3.

2.6. Statistical analysis

The obtained results (both manual and IC measurements) were entered into a database and imported in the statistical software R (http://cran.r-project.org/). Visual comparison of the two methods was performed by Bland-Altman (BA) plots [13] of un-transformed and 4th root transformed values to obtain approximate variance homogeneity (Supplementary Fig. S1). Analytical comparison of the two methods was performed with standard regression methods (least square regression) as well as with non-parametric Passing-Bablok regression (Passing and Bablok, 1983, estimated with the "PaBaLarge" function and 10^6 bins within the 'mcr' R package) which assumes measurement error in both methods and is in accordance with the Clinical and Laboratory Standards Institute recommendations for analytical method comparison and bias estimation using patient samples (EP09-A3, August 2013). The limits of agreement (LoA, Table 1) were defined as mean ± 2 times the standard deviation of the difference between measurements. Wilcox-rank sum test was used to test for differences between groups of differences in repeated measurements. The coefficient of variation (CV) was defined as the standard deviation divided by the mean and expressed as a percentage.

2.7. Definition of outliers

Some assessments clearly deviated when comparing the two methods. Most likely, these outliers were caused by errors in sample handling or preparation of dilutions etc. In order to remove these

Table 1
Summary statistics for the data.

	All samples	All ex. outliers & azo	Interval[a] [0–3]	Interval[a] [3–12]	Interval[a] [12–200]
Number of samples	4010	3705	657	625	2088
Mean manual	32.64	35.07	0.59	6.77	54.64
Mean IC	32.78	35.18	0.99	7.57	54.48
Median manual	15.78	18.63	0.4	6.41	43.36
Median IC	15.62	18.58	0.69	7.51	42.3
PB slope	0.996 (0.989–1.004)	0.987 (0.979–0.994)	1.578 (1.5–1.66)	0.901 (0.863–0.947)	1 (0.988–1.012)
PB intercept	0.192 (0.15–0.246)	0.414 (0.349–0.492)	0.034 (0.027–0.044)	1.33 (1.049–1.576)	0.35 (−0.691–0.062)
Ordinary regression slope	0.971 (0.957–0.985)	0.973 (0.958–0.986)	1.241 (1.17–1.312)	0.691 (0.644–0.731)	0.979 (0.962–0.996)
Ordinary regression intercept	1.092 (0.764–1.444)	1.058 (0.719–1.384)	0.253 (0.215–0.291)	2.9 (2.601–3.268)	0.983 (0.294–1.772)
Residual standard error	8.491	7.719	0.472	1.467	9.383
Mean of difference (bias)	0.133 (−0.133–0.398)	0.104 (−0.147–0.356)	0.541 (0.491–0.591)	0.949 (0.813–1.086)	−0.196 (−0.594–0.202)
Upper LoA	17.3 (16.92–17.68)	15.72 (15.37–16.08)	1.93 (1.86–2)	4.79 (4.59–4.98)	18.7 (18.14–19.27)
Lower LoA	−17.03 (−17.41–−16.66)	−15.51 (−15.87–−15.16)	−0.85 (−0.92–−0.78)	−2.89 (−3.08–−2.7)	−19.1 (−19.66–−18.53)
Normalized bias[b]	0.004	0.003	0.54	0.132	−0.004
Normalized residual standard error[b]	0.26	0.22	0.471	0.205	0.175
PB slope derivation from 1	0.004	0.013	0.578	0.099	0
Pearsons R	0.98	0.984	0.851	0.766	0.97

PB: Passing Bablok, LoA: limits of agreement,
[a] Samples being within the given interval for both methods.
[b] The residual standard error, bias or LoA divided by the mean of the interval.