Wales Fertility Institute

UNCERTAINTY OF MEASUREMENT

1 WHAT IS UNCERTAINTY?

There several definitions of uncertainty:
"A parameter associated with the result of measurement that characterizes the dispersion of the values that could be reasonably attributed to the measurand"[1]
"Non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used"[2]

Uncertainty of measurement comprises of several components, some of which can be measured and characterised by standard deviation, others can be evaluated from assumed probability distributions based on experience or other information.

2 EXPRESSING UNCERTAINTY OF MEASUREMENT

Uncertainty relates in general to the concept of doubt. In this document it refers to either a parameter associated with the definitions above, or to limited knowledge about a particular value. Uncertainty of measurement does not imply doubt about the validity of a measurement; on the contrary, knowledge of uncertainty implies increased confidence in the validity of a measurement value.

It is important to realise that uncertainty is not the same as error. Error is the difference between the measured value and the 'true value' of the thing being measured, whereas uncertainty is the quantification of the doubt about the measurement result. Any error whose value we do not know is a source of uncertainty.

Whilst it is possible to quantify the measurement uncertainty of some of the individual quantitative elements of a semen analysis, it is not possible to derive the uncertainty of measurement of the semen analysis result as a whole due to several unavoidable and uncontrollable confounding factors described below.

3 WHY IS IT IMPORTANT TO CONSIDER UNCERTAINTY?

Uncertainty is a (usually quantitative) indication of the quality of the result. It gives an answer to the question, how well does the result represent the value of the quantity being measured? It allows users of the result to assess its reliability, for example for the purposes of comparison of results from different sources or with reference values or ranges.

In the case of a semen analysis, a result is often compared to a reference range. In this case, knowledge of the uncertainty shows whether the result is well within the reference range or only just makes it. Sometimes a result is so close to the limits of the reference range that the risk associated with the possibility that the measured parameter may not fall within the limit, once the uncertainty

has been allowed for, must be considered. In other words, results which fall just outside the normal range may in reality be within the normal range.

We often encounter patients who get differing results (and therefore sometimes conflicting advice) when they have semen analyses performed in more than one laboratory. Whilst this is often due to the inherent within-individual variability of semen samples a further complication may be arising. For example, if a patient were able to have the same sample analysed simultaneously in two Andrology labs. Would we expect the laboratories to get identical results? Only within limits, we may answer, but when the results are close to the specification limit it may be that one laboratory indicates 'normality' whereas another indicates an 'abnormality'. From time to time accreditation bodies have to investigate complaints concerning such differences. This can involve much time and effort for all parties, which in many cases could have been avoided if the uncertainty of the result had been known by the service user.

4 WHERE DO UNCERTAINTIES IN SEMEN ANALYSIS COME FROM?

Many things can undermine a measurement of a semen analysis parameter and importantly these flaws in the measurement may be visible or invisible. Although patients and Andrologists do their best, the nature of semen analyses dictate that they are rarely performed under absolutely perfect conditions and as such, errors and uncertainties can arise from the areas detailed in the table below. However, in some areas it is possible to attempt to control for and minimise these errors and uncertainties and the ways in which we attempt to do this is also included below.

Source of error or uncertainty	**Control/minimization methods**
The laboratory equipment used to perform measurements. Measuring instruments (pipettes, counting chambers etc.) can suffer from errors including bias, changes due to ageing, wear, or other kinds of drift, poor readability, noise (for electrical instruments) and many other problems. Please note that the uncertainty of measurement generated by the pipettes used during semen analysis small compared with the other sources of uncertainty described in sections below	• Formal installation and validation • Regular maintenance • Calibration

The patient It is well recognized that the 'quality' of semen samples that a man produces can vary hugely for a variety of reasons, not least of which is normal biological variation. As such a diagnosis should not be based on only one semen analysis.	• Performance of repeat semen analyses to help derive a 'representative' diagnosis
The semen sample itself Human semen is a heterogeneous fluid which undergoes a process of liquefaction shortly after ejaculation. The constituents of seminal plasma are not capable of sustaining sperm motility and viability over prolonged periods	• The Andrology Laboratory examines the sample within 60 minutes of it being produced wherever possible • Semen samples are well mixed before aliquots are removed for assessment purposes • Awareness that sampling a non-liquefied sample may lead to an erroneous result
Semen sample collection The way in which a semen sample is collected can hugely affect its quality. • Duration of abstinence • Collection method • Collection vessel • Incomplete collection • Exposure to adverse temperature • Ejaculation to analysis interval	• Patients are advised to abstain from ejaculation for a minimum of two and a maximum of seven days. • Patients are advised to collect their samples by masturbation • Patients are advised to only use the container provided by the Andrology Laboratory • Patients are advised to inform the Andrology Laboratory if any of the sample was spilled. • Patients are advised to protect the sample from extremes of temperature • Patients who produce their sample off-site are advised to deliver the sample to the Andrology laboratory within 60 minutes of it being produced

Imported uncertainties Calibration of for example, pipettes or heated-stages will have an inherent uncertainty which is then built into the uncertainty the measurement being made. NB. The uncertainty due to not calibrating equipment would be significantly more.	• It is not possible to control for this *per se* although it is essential that equipment is regularly calibrated so that the level of uncertainty is known.
Operator skill and judgment Some measurements (e.g. assessment of sperm motility by eye) depend upon the skill and judgment of the person looking down the microscope. For example a sperm is deemed to be progressing rapidly (i.e. grade A) if it is moving >25µm/sec which equates approximately to 5 x the length of a sperm. Such an assessment is highly subjective. Similarly, the human eye is unavoidably drawn to moving objects and as such is inclined to overestimate sperm motility	• Training • IQC • EQA • Use of computer-aided semen analysis (CASA)
Sampling issues The measurements that are made relating to a particular semen sample must be properly representative of the semen sample itself. This is particularly relevant as human semen is a heterogeneous fluid which undergoes a process of liquefaction shortly after ejaculation.	• Semen samples are well mixed before aliquots are removed for assessment purposes • Awareness that sampling a non-liquefied sample may lead to an erroneous result
The environment Temperature, air pressure, humidity and many other conditions can affect the measuring instrument or indeed, the sample being measured	• Patients are advised to protect the sample from extremes of temperature • Patients are advised to only use the container provided by the Andrology Laboratory • All motility assessments are performed at 36^{o}C ± 2^{o}C

So, it can be seen from the table above that there are many very real issues which may cause uncertainty of measurement in relation to a semen analysis and although we can do our best to control for these, many of the control methods listed above rely heavily on patient compliance.

5 QUANTIFICATION OF UNCERTAINTY

This section seeks, where possible, to quantify the uncertainty of measurement for individual parameters within a semen sample. This allows individual parameters within the analysis to be appropriately interpreted. The laboratory has where possible attempted to determine the uncertainty relating to each parameter.

The all staff within the laboratory assesses a number of different samples, attempting to cover the expected ranges of patient samples, for which the “target value” had been determined. For each member of staff and sample, the percentage deviation from the target was then calculated. From this the means and standard deviations could be calculated. Assuming a normal distribution of results about the actual value, the laboratory accepts that to ensure there is 95% coverage of the actual result, the uncertainty about the reported result is ±2 standard deviations. Within this document the uncertainty of the reported value will be expressed as ±xx%

At the end of each section, we have attempted to relate the data to clinical practice by suggesting some **points for consideration**, which may help your interpretation.

5.1 Within-patient variation

It is well recognised that an individual patient will show considerable variability of results between individual semen samples.

Points for consideration - clearly, men will produce samples of very variable quality. Diagnosis of sub-normality should not be based on a single semen sample.

5.2 Volume measurement

Semen volume is measured by weight. Uncertainty related to the measurement of semen sample volume is very small as demonstrated:

10 replicate measurements of a standard weight (Cardiff).

1	2	3	4	5	6	7	8	9	10		Mean	SD
14.7433	14.7433	14.7431	14.7433	14.7433	14.7433	14.7433	14.7432	14.7431	14.7431		14.7432	0.0001

10 replicate measurements of a standard weight (Singleton).

1	2	3	4	5	6	7	8	9	10		Mean	SD
14.96	14.97	14.96	14.96	14.97	14.97	14.97	14.96	14.97	14.97		14.965	0.0053

Calibration data has demonstrated that the maximum uncertainty of measurement of the balance is ±0.002%. This is for the stated range of measurements of 0 – 200g.

Points for consideration – the reported volume of a semen sample is extremely reliable.

5.3 Concentration

The laboratory has 3 pipettes that are used to prepare the dilutions required for the manual assessment of concentration; these have been calibrated by a UKAS accredited calibration laboratory.

Cardiff

Pipette	Maximum inaccuracy (%)	Maximum imprecision (%)	Maximum uncertainty (µl)
Gilson Microman 10-100µl (HB05049)	**1.05**	**0.33**	**0.83**
Finnpipette Classic 40-200µl (2600)	**1.39**	**0.28**	**0.87**
Finnpipette Classic 200-1000µl (2406)	**7.68**	**0.33**	**4.04**

Maximum combined uncertainty for pipetting when creating dilution:
$\sqrt{0.87^2 + 4.04^2} = 4.133$µl

Singleton

Pipette	Maximum inaccuracy (%)	Maximum imprecision (%)	Maximum uncertainty (µl)
Gilson Microman 10-100µl (LJ05048)	**0.45**	**0.19**	**0.71**
Genex 10-100µl (8509705)	**0.49**	**0.16**	**0.66**
Sealpette Pro 100-1000µl (5712319)	**7.93**	**0.65**	**4.21**

Maximum combined uncertainty for pipetting when creating dilution:
$\sqrt{0.66^2 + 4.21^2} = 4.261$µl

For the manual determination of concentration, the laboratory has several Improved Neubauer Haemocytometers. These are individually identified, by year in use and a sequential number. Twice a year all the haemocytometers are calibrated using QC-Beads (Microm). Any haemocytometer which is found to be significantly different from the others should be discarded.

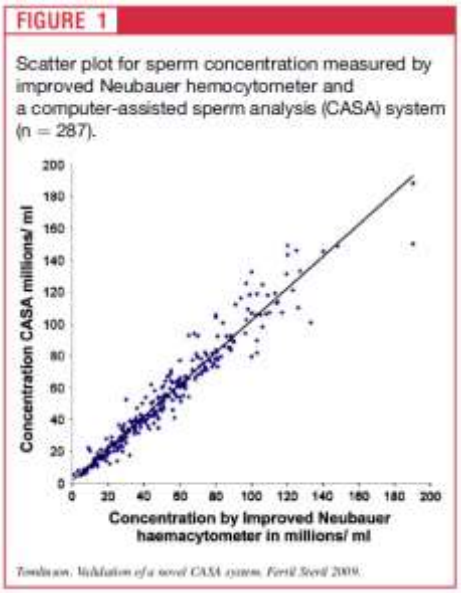
Cardiff

QC-Beads	Expected range ($x10^6$/ml)	Mean ($x10^6$/ml)	SD	Min ($x10^6$/ml)	Max ($x10^6$/ml)
Lo	16 – 24	16.51	0.93	15.00	18.18
Hi	34 – 46	30.68	1.84	27.38	33.42

Singleton

QC-Beads	Expected range ($x10^6$/ml)	Mean ($x10^6$/ml)	SD	Min ($x10^6$/ml)	Max ($x10^6$/ml)
Lo	16 – 24	16.41	1.61	13.7	19.6
Hi	34 – 46	32.31	2.72	28.0	37.4

Figure 1 below shows the correlation between CASA and 'manual' sperm concentration measurements for 287 semen samples.



These data clearly show that the Sperminator CASA machine can measure sperm concentration at least as well as the conventional haemocytometer method.

Points for consideration – whenever possible concentration measurements are performed using a CASA machine.

5.3.1 Measurement of sperm concentration using CASA

Although we use CASA machines wherever possible measurement uncertainty still exists.

Variability observed when using CASA to measure sperm concentration on 10 aliquots of the same sample.

n=10	
Mean concentration (millions/ml)	**13.42**
Range	**11.2-15.8**
SD	**1.48**
SEM	**0.47**
CV	**0.11**

Variability observed when using CASA to measure sperm concentration on 5 measurements of the same aliquot of the sample

n=5	
Mean concentration (millions/ml)	**18.74**
Range	**17.94-19.00**
SD	**0.44**
SEM	**0.20**
CV	**0.02**

Repeated measurement of the same aliquot, demonstrates a lower CV than multiple aliquots of the same. This highlights the inherent heterogeneity of human semen.

Points for consideration – the uncertainty associated with measuring sperm concentration, even using CASA technology is quite high.

5.3.2 Manual measurement of sperm concentration

Unfortunately, CASA systems require reasonable numbers of sperm present in the sample (more than approximately 5 million per ml) to function optimally. Therefore, we remain reliant on manual measurement of sperm concentration for semen samples containing lower concentrations of sperm. As such, it is prudent to examine the uncertainty associated with the manual measurements of sperm concentration.

5.3.2.1 Within-observer variability

The table below shows the results of 10 manual concentration measurements performed on the same sample by the same operator.

Variability observed with the same observer performing a manual concentration measurement on the same sample10 times

n=10	
Mean concentration (millions/ml)	**36.76**
Range	**34.63-39.67**
SD	**2.12**
SEM	**0.67**
CV	**0.06**

The CV is reasonable and similar to that those derived using the CASA system.

5.3.2.2 Between-observer variability

The table below shows the results of 8 operators each performing manual concentration measurements on 4 different samples. Please note that the variability seen here may be a combination of true 'between-observer' variability together with sampling error.

Variability observed with 8 observers performing a manual concentration measurement on the same sample at the same time

Sample Nº.	A	B	C	D	E	F	G	H	Mean	Max	Min	SD	CV
1	**15.86**	**10.73**	**9.05**	**13.90**	**12.40**	**13.31**	**12.08**	**10.85**	**12.33**	**15.86**	**9.05**	**2.03**	**0.16**
2	**49.10**	**51.38**	**49.38**	**65.40**	**66.70**	**44.67**	**41.33**	**55.13**	**59.63**	**66.70**	**41.33**	**7.41**	**0.12**
3	**23.75**	**25.25**	**20.50**	**22.70**	**20.00**	**20.98**	**21.60**	**25.63**	**22.55**	**25.63**	**20.00**	**2.15**	**0.10**
4	**53.70**	**43.08**	**51.13**	**60.10**	**52.50**	**46.08**	**54.25**	**67.17**	**53.50**	**67.17**	**43.08**	**7.57**	**0.14**
												Mean CV	**0.13**

Once again, the CVs are large and considerably more than the CASA, particularly when concentration is lower.

Anova: Single Factor - Concentration						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Target	28	1208.82	43.17214	521.878		
Operator 1	28	1202.65	42.95179	508.7415		
Operator 2	28	1195.45	42.69464	567.8272		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3.198402	2	1.599201	0.003001	0.997003	3.109311
Within Groups	43158.06	81	532.8156			
Total	43161.26	83				

Analysis of variance shows good correlation between operators and target values

Uncertainty associated with manual concentration = ±42.83%

Points for consideration – the uncertainty associated with manually measuring sperm concentration can be large particularly at lower sperm concentrations.

5.4 Motility

There are principally four areas in which uncertainty of measurement can be introduced when measuring sperm motility these being

i. the time interval between ejaculation and analysis
ii. the effect of temperature
iii. the effect of the operator or CASA system and
iv. the difference between operators

5.4.1 Time interval between ejaculation and analysis

Sperm motility in some semen samples will start to decline after approximately 60 minutes. As such, the Andrology Laboratory endeavors to perform all motility analyses within 60 minutes of ejaculation.

Points for consideration –Patients who produce samples off-site should be strongly advised to deliver the samples to the Andrology Lab within 60 minutes of ejaculation.

Any instance where sperm motility assessment was performed over 70 minutes after ejaculation (e.g. where the sample was produced off-site) will be highlighted on the report.

5.4.2 Effect of temperature

To ensure that all assessment of sperm motility is comparable and not subject to variation in environmental conditions, all assessments are performed using a heated stage.

Points for consideration - All motility analyses performed within the WFI - Andrology Laboratories are performed at 36^{o}C ± 2^{o}C.

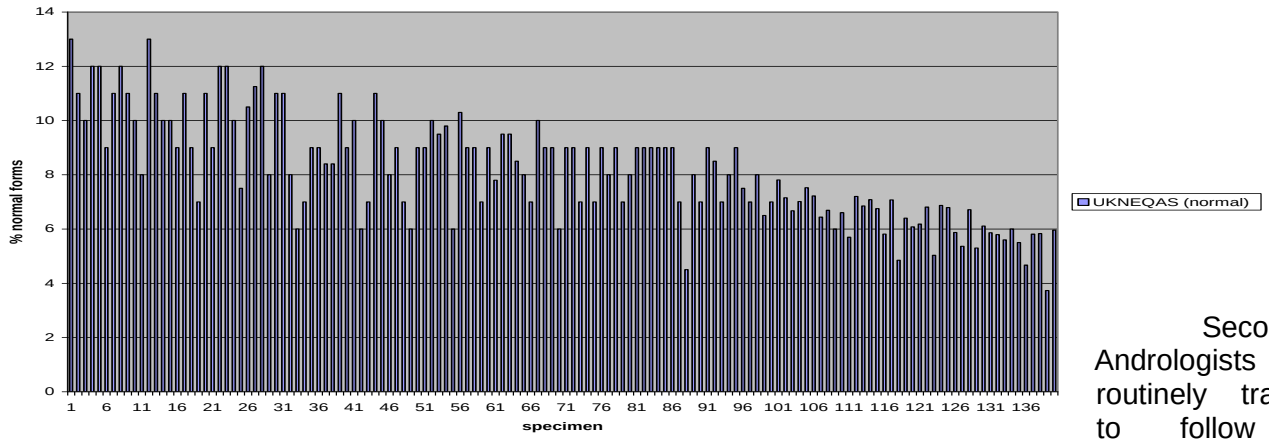
Failure by patients to follow instructions regarding sample production and transport will be noted on the report.

5.5 Morphology

The assessment of sperm morphology is fraught with difficulty for many reasons and significant measurement uncertainty exists. Some examples of these difficulties are given below.

The figure below would suggest that a laboratories' perception of 'a normal sperm' is slowly changing to meet the needs of the new reference ranges, despite using the same sperm shape and size criteria to work to. Figure 1. below show the target values for % normal forms from EQA samples over the past 8 years. There is a clear relationship showing generally stricter scoring with time in response to a gradual adoption of a lower reference range.

EQA target values for % normal forms from 2005-2011 in the UK NEQAS scheme

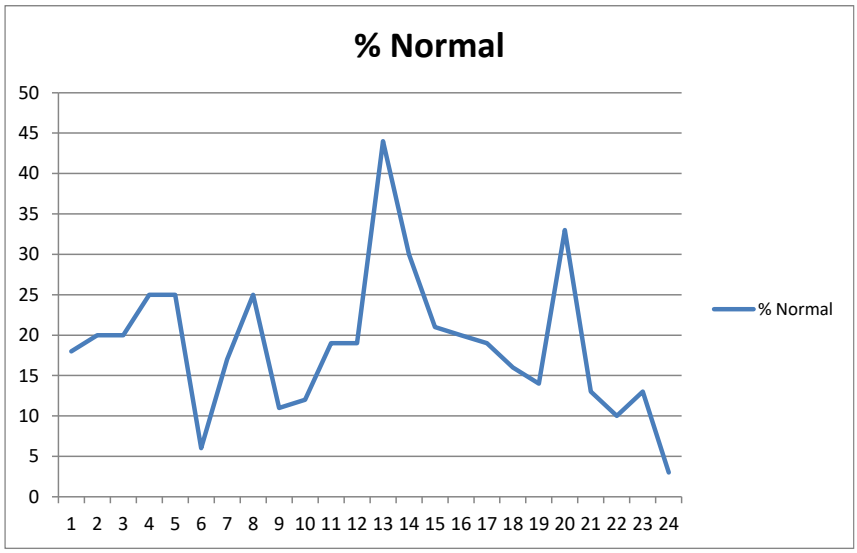


Secondly, Andrologists are routinely trained to follow a philosophy which labels any sperm which does not meet pre-defined size and shape definitions as being abnormal. By definition the group of abnormal forms then includes a significant number of 'unknowns' which could include: borderline forms; artefacts created by slide preparation; or indeed those which become adhered to artefacts such as debris or non-sperm cells. The consequence of adopting this strategy is that the uncertainty surrounding those 'unknowns' (and therefore for the entire measurement) cannot be assessed.

Thirdly, and to compound the difficulties yet further, not only can differences in fixation and staining make a difference to the overall result but individual interpretation of exactly the same sperm images show a remarkable lack of consistency across a range of operators.

A recent small study using a series of clear images sent out by the laboratory at Nottingham University Hospital to a number of centres showed that even in experienced hands, agreement on whether a sperm is normal or abnormal varied considerably. The figure below shows the % normal forms as reported by 24 individuals (fully trained to perform semen analysis) based in six different laboratories. The mean from 160 sperm images was 18.9% normal with a range from 3% to 44%!!!

Identification of normal forms from the same set of micrographs assessed by 24 staff in six different centres



Clearly there is a very large and unquantifiable uncertainty associated with sperm morphology assessment as currently performed, and it seems that estimating sperm morphology in terms of percentage 'normal forms' is difficult (if not impossible) with subjectivity remaining a significant problem.

However, there are certain situations where the performance of sperm morphology is associated with an extremely low (if not zero) level of uncertainty and these are where the morphological defect applies to every sperm and such conditions are easily recognisable. This might include conditions such as globozoospermia (where the head size is increased and no acrosome is present), pin-head sperm (where the sperm heads are missing) or gross tail defects. Such conditions are often 'sterilising'.

5.6 Estimates of uncertainty associated with diagnostic semen analysis

All staff within both laboratories regularly take part in both IQC and EQA, this provides the laboratory with an ongoing source of data to determine intra-operator variability and therefore uncertainty.

	WHO 2010 lower ref limit	WHO 2010 95% CI	UoM (Mean deviation)	UoM (%)	UoM applied to lower ref limit
Concentration (Manual)	15	12 – 16	-0.29	1.93	14.71
Concentration (CASA)	15	12 - 16	2.16	14.4	12.84
Motility (PR)	32	31 – 34	1.91	5.97	30.09
Vitality	58	55 – 63	0.36	0.62	57.64
Morphology	4	3 – 4	-0.94	23.5	3.06
Round cells	5 (guide value)	N/A	-0.87	17.4	4.13

5.7 Post-vasectomy semen analysis

PVSA following a vasectomy is to determine clearance of sperm and also to establish either technical surgical failure or early recanalization. The WHO 2010 manual[3] state that "only when no sperm

spermatozoa are found after a complete and systematic search of all the resuspended precipitate should samples be classified as azoospermic". In practice this is a description, "no sperm seem". The analysis requires is to determine a result that is below the limit of detection of the method. If a single sperm is seen it is likely to be below the statistically derived limit of detection. It is impossible to determine azoospermia with statistical probability.

For the large-volume, fixed-depth, disposable counting chamber method[4] used for PVSA by the laboratory, the current statistically derived limit of detection is 244 sperm/ml

Points for consideration – Determination of azoospermia is statistically impossible. Where the laboratory reports "No sperm seen", this means that there may be sperm present below the limit of detection. When giving the patient clearance following their operation the clinician should be aware that very small numbers of sperm may be present.

6 SUMMARY OF CONSIDERATIONS & ADVICE

For your convenience, we have provided below a summary which seeks to draw together the main points when considering measurement uncertainty and semen analysis.

- **It is essential that patients be strongly advised to follow instructions regarding sample collection and abstinence to reduce the uncertainty that this can introduce.**
- **An interval of more than an hour between ejaculation and analysis may lead to a reduction in sperm motility – this will be highlighted on the report.**
- **Results from samples which are not fully liquefied may not be truly representative of the sample's quality – this will be highlighted on the report.**
- **Men will produce samples of very variable quality due to normal biological variation. As such, a diagnosis of sub-normality should not be made on a single semen sample.**
- **The measurement of sperm concentration (either manually or by computer) is associated with a high degree of measurement uncertainty and this should be taken into account when interpreting semen analysis results, particularly at the limits of normality.**
- **The measurement of sperm motility (either manually or by computer) is associated with a very high degree of measurement uncertainty and this should be taken into account when interpreting semen analysis results, particularly at the limits of normality.**
- **The measurement of sperm morphology is associated with a very high degree of measurement uncertainty and this should be taken into account when interpreting semen analysis results. Measurement of sperm morphology is of considerable value in identifying gross morphological abnormalities.**
- **PVSA is not a technique by which azoospermia can be determined. When "No sperm seen" is reported, it is implied that there are no sperm present that can be detected with a high degree of statistical probability.**

7 CONCLUSIONS

As a user of the diagnostic andrology service – what do I need to do?

The simple answer to this question is nothing. The way in which semen analysis testing is performed and the inherent problems and difficulties therein remain unchanged. Similarly, the uncertainty of measurement associated with performing a routine semen analysis has, and will always be present to a greater or lesser degree.

As the provider of your semen analysis testing we would simply ask, having taken the time to read this document, that you consider its content when interpreting a semen analysis result within the clinical environment.

8 REFERENCES

1. Guide to the Expression of Uncertainty in Measurement. ISO, Geneva (1993). (ISBN 92-67-10188-9) (Reprinted 1995: Reissued as ISO Guide 98-3 (2008))
2. ISO/IEC Guide 99:2007, International vocabulary of metrology – Basic and general concepts and associated terms (VIM). ISO, Geneva, (2007).
3. Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interaction. 5th Ed. WHO, Cambridge, Switzerland (2010).
4. Hancock P et al. (2014). Use of large-volume, fixed-depth, disposable slides for post-vasectomy semen analysis. *British Journal of Biomedical Science*, **71**(1):1-5.