NATIONAL INSTITUTE OF PUBLIC HEALTH AND THE ENVIRONMENT

BILTHOVEN

Report no. 650030.001

Is Sperm Quality Actually Declining ?
A Literature Survey

M.M. Mees, C.E.J. Cuijpers, A.H. Piersma

November 1997

Mrs.M.M.Mees has performed this study as part of the requirements for her study Biology at the University of Leiden.

This study was performed in order and for the account of the Ministries VWS and VROM within the framework of the project "Sperm Quality and Estrogenic Environmental Contaminants" (project no. 650030).

National Institute of Public Health and the Environment, PO Box 1, 3720 BA Bilthoven, the Netherlands, tel.: .. 31 30 274 91 11; fax.: .. 31 30 274 29 71

VERZENDLIJST

01 - 05	Directie Gezondheidsbeleid, Mr S. van Hoogstraten
06 - 10	DGM, Directie SVS, Dr C.M. Plug
11	Directeur-Generaal Dr H.J. Schneider
12	Plv. Directeur-Generaal Milieubeheer, Dr Ir B.C.J. Zoeteman
13	Prof. J.J. Sixma, voorzitter van de Gezondheidsraad
14	Dr J.A.M. Hulshof, VWS/GZB
15	Dr J.J. Ende, VWS/GZB/CO
16	Dr H. Roelfzema, VWS/GZB/CO
17	Drs G.E.H. Hoeben, VWS/GZB/CO
18	Dr W.H. van Eck, VWS/GZB/VVB
19	Drs N.B. Lucas Luijckx, VWS/GZB/VVB
20	Drs J.J.L. Pieters, VWS, Hoofdinspectie Gezondheidszorg
21	Dr Ir P.C. Bragt, VWS, Hoofdinspectie Gezondheidsbescherming
22	Dr Ir G. Kleter, VWS, Hoofdinspectie Gezondheidsbescherming
23	Ir M. Bovenkerk, VROM, DGM/SVS
24	Dr J. van Zorge, VROM/SVS
25	Dr ME.J. van der Weijden, VROM/SVS
26	Dr M.N.E.J.G. Philippens, VROM, DGM/SVS
27	Dr R.J. van Wijk, AKZO/Nobel, afd. RGL, Arnhem
28	Prof. Dr J.G. Koppe, AMC/afd. neonatologie, Amsterdam
29	Dhr. M. Tonkes, Aquasense, Amsterdam
30	Dr B. Bosveld, DLO-IBN, Wageningen
31	Dr J. Stab, DZH, Voorburg
32	Prof. Dr N. van Straalen, VU/fac. Biologie, Amsterdam
33	Dr W.F. Passchier, Gezondheidsraad, Rijswijk
34	Dr J.W. Dogger, Gezondheidsraad, Rijswijk
35	Mevr. A. van der Ven, Greenpaece NL, Amsterdam
36	Mevr. H.B. Oonk, ID/DLO/afd. Moleculaire herkenning, Lelystad
37	Dhr. D. Meijer, Inspectie Gezondheidsbescherming, Zutphen
38	Dr A. Belfroid, IVM, Amsterdam
39	Dr H. Jenner, KEMA-KES, Arnhem
40	Dr W. Denneman, KIWA, Nieuwegein
41	Prof. Dr Wendelaar-Wonga, KUN/vakgr. Toxicologie, Nijmegen
42	Prof. Dr W. Koeman, LUW/vakgr. Toxicologie, Wageningen
43	Prof. Dr J. Kleinjans, RU Limburg, Maastricht
44	Vereniging Milieudefensie, Amsterdam
45	Mevr. Drs M. Klein, NBLF-IKC, Wageningen
46	Dr P.C. Nordam, SoZaWe
47	Drs J. Marquenie, NAM, Assen
48	Mevr. M. Addink, Natuurhistorisch Museum, Leiden
49	Dhr. Beurend, Ned. Ver. Zeepfrabikanten, Zeist
50	Dr J. Everaarts, NIOZ, Den Burg (Texel)
51	Dhr. J. de Graaf, Organon Nederland, Oss
52	Ir LG.M.Th. Tuinstra, RIKILT, Wageningen
53	Dr D. Verhaak, RIKZ, Middelburg
54	Dr M. van den Berg, RITOX, Utrecht

55 Prof. Dr W. Seinen, RITOX, Utrecht
56 Dr R.A.C. Lock, KUN/vakgr. Dierfysiologie, Nijmegen
57 Prof. Dr P. van der Saag, Hubrechtlaboratorium, Utrecht
58 Prof. Stegeda, Stichting C3, Amsterdam
59 Dr P. Reijnders, DLO-IBN, Den Burg (Texel)
60 Dr A. Brouwer, LUW/vakgr. Toxicologie, Wageningen
61 Dr W. de Wolf, NV Proctor & Gamble, ETC, Strombeek-Bever, Belgie
62 Dr J. Boon, NIOZ, Den Burg (Texel)
63 Dr T. van Brummelen, RWS/Directie Noordzee, Rijswijk
64 Dr A. Gerritsen, TNO, Delft
65 Dr Ran, Academisch Ziekenhuis VU, Amsterdam
66 Dr A. Rijnsdorp, RIVO, IJmuiden
67 Dr J. Hendriks, RIZA, Lelystad
68 Prof. Dr H. Goos, RUU/vakgr. Experimentele Dierkunde, Utrecht
69 Dr J. Lambert, RUU/vakgr. Experimentele Dierkunde, Utrecht
70 Dr J. Parsons, UVA/MTC (Sense), Amsterdam
71 Prof. Dr L. Reijnders, UVA/Milieukunde, Amsterdam
72 Dr Ir N. Roeleveld, KUN/vakgr. Medische Informatiekunde, Epidemiologie en Statistiek, Nijmegen
73 H. Thuis, KUN/vakgr. Medische Informatiekunde, Epidemiologie en Statistiek, Nijmegen
74 Dr W. de Kort, TNO Voeding, div. Arbeidstoxicologie en voeding, Zeist
75 Dr D. Heederik, LWU/vakgr. Humane Epidemiologie en Gezondheidsleer, Wageningen
76 Dr Ir F.J. Jongeneelen, Industox Consult, Nijmegen
77 Dr J.H. van Wijnen, GG en GD, Amsterdam
78 Dr R. Luijk, Consumentenbond, Den Haag
79 Drs M. Lursen, Wetenschapswinkel Biologie Univers. Utrecht, Utrecht
80 Dr R.F.A. Weber, Erasmus Universiteit, Rotterdam
81 Dhr. Zaadstra, TNO-Preventie, Leiden
82 Depot Nederlandse Publikaties en Nederlandse Bibliografie
83 Directie RIVM
84 - 85 SBD/Voorlichting & Public Relations
86 Drs C. van der Heijden, WHO, Bilthoven
87 Dr F.X.R. van Leeuwen, WHO, Bilthoven
88 Dr Ir G. de Mik, Directeur Sector 4
89 De leden van de RAG
90 Dr A. Opperhuizen, hoofd LEO
91 Dr W.H. Konemann, hoofd CSR
92 Ir J.J.G. Kliest, hoofd IEM
93 Drs J.W. Dorpema, hoofd LGM
94 Dr H.J.P. Eijsackers, hoofd ECO
95 Dr J.G. Vos, hoofd LPI
96 Dr J. Meulenbelt, hoofd NVIC
97 Dr Ir H.J.G.M. Derks, hoofd LGO
98 Mevr. M.J. Garbis-Berkvens. NVIC
99 Prof. Dr P.W.J. Peters, SBD

100	Dr P. Kramers, VTV
101	Dr P. Wester, LPI
102	Dr R. Leewis, LWD
103	Dr W. Slooff, MNV
104	Drs M. Vaal, ECO
105	Drs J.H. Canton, ECO
106	Dr C.E.J. Cuijpers, CCM
107	Dr E. Lebret, CCM
108	Drs A.E.M. de Hollander, CCM
109	Prof. Dr C. van Leeuwen, CSR
110	Drs A.G.A.C. Knaap, CSR
111	Dr Ir J.G.M. van Engelen, CSR
112	Dr W.C. Mennes, CSR
113	Dr E.J. de Waal, LGM
114	Mevr. Dr A.J.A.M. Sips, LBO
115	Drs A.K.D. Liem, LOC
116	Dr L. van Bree, LEO
117	Dr E.H.J.M. Jansen, LEO
118	Dr C.F. van Kreijl, LEO
119	Dr ir M.N. Pieters, LEO
120 - 122	Auteurs
123	Bibiliotheek RIVM
124	Bibliotheek CSR
125	Bureau Rapportenregistratie
126 - 145	Bureau Rapportenbeheer

CONTENTS

SUMMARY

Recently, a possible decline in human semen quality has become a major issue of concern. In 1992 a meta-analysis by Carlsen and colleages claiming a 50% reduction in semen quality over the last 50 years was published, which created quite a stir in the media. However, the supposed decline in human sperm quality was far from unambiguous since this publication received much criticism and other publications on a time trend showed contradictory results.
This study was designed to assess the likelihood of a true decline, through a critical analysis of the meta-analysis of Carlsen *et al.* and epidemiological publications that describe a trend in sperm parameters. The design and performance of each of the separate studies was evaluated with regard to possible bias and confounders, methods of sperm measurement and statistics, to estimate the validity of their results.
The qualitative analysis of the meta-analysis by Carlsen *et al.* (1992) as well as the primary epidemiological studies revealed that the evidence to justify the conclusion that sperm quality or quantity has declined over the past decades is not convincing.

SAMENVATTING

De kwaliteit van menselijk zaad is recent ter discussie gekomen naar aanleiding van publicaties die wezen op een teruggang in sperma-aantallen in de tijd. In 1992 verscheen een meta-analyse van epidemiologische studies, waarin Carlsen en collega's een 50% reductie in sperma-aantallen over de laatste 50 jaar beschreven. Deze studie resulteerde in uitgebreide berichtgeving in de media. De gesuggereerde reductie werd in de wetenschappelijk literatuur sterk bekritiseerd, bovendien vertoonden andere studies op dit gebied tegenstrijdige resultaten.
Dit onderzoek werd uitgevoerd om de waarschijnlijkheid van een echte reductie in spermakwaliteit vast te stellen, door een kritische analyse van de studie door Carlsen en andere studies op dit gebied. De opzet en uitvoering van elk van de studies werd geëvalueerd met het oog op bronnen van bias en confounders, methoden van sperma-analyse en statistiek, teneinde de validiteit van de resultaten te beoordelen.
De kwalitatieve analyses van zowel de meta-analyse van Carlsen als de primaire epidemiologische studies gaven aan dat de onderbouwing van de conclusie dat de kwaliteit en/of kwantiteit van het menselijk sperma de laatste decennia is afgenomen niet overtuigend is.

1. INTRODUCTION

During the last five years, a possible decline in human semen quality has become a major issue of concern. In 1992 an article was published claiming a 50% reduction in human sperm quality over the last 50 years (Carlsen *et al.* 1992). The authors interpreted this as an indication of declining male fertility. A common aetiology was supposed with the increased incidence of other male reproductive disorders such as testicular cancer, hypospadias and cryptorchidism (Adami *et al.* 1994; Matlai & Beral 1985; Ansell *et al.* 1992), for which the evidence seems stronger (Sharpe 1993). As a possible explanation for their findings they suggested the antenatal action of environmental estrogens. Findings regarding the treatment of pregnant women with DES-hormone between the late 1940s and late 1970s were used to support this explanation. Follow-up of the in utero exposed children learned that abnormalities in the reproductive organs were more frequent than in non-exposed children (Henderson *et al.* 1976; Gill *et al.* 1979; Stillman 1982). However, the therapeutical dose of DES was very high as compared to usual exposure to estrogens from food or the environment.

Meanwhile many publications concerning the effects and presumed harmfulness of environmental estrogens on wildlife have appeared. Several of the phenomena observed in wildlife, such as a decreasing fertility in birds, gastropods and mammals, reduced reproductive success in birds, fish and turtles, aberrant thyroid function in birds and fish, and sex-reversal in fish (review in Toppari *et al.* 1995) could be attributed to the detrimental action of xeno-estrogenic compounds (Purdom *et al.* 1994; Bergeron *et al.* 1994).

At the same time, the popular media have discovered the subject. Especially the BBC Horizon documentary 'Assault on the male' (1993) and the recently published book 'Our stolen future' (Colborn *et al.* 1996) provided great publicity to the possible consequences of exposure to xeno-estrogens. This, and the regularly appearing newspaper articles commenting on new scientific findings (Parool 1996; Nederlands Dagblad 1996 and others), resulted in considerable public turmoil.

Being put forward as bright and clear in the popular media, the issue of environmental estrogens and sperm quality in scientific literature is far from unambiguous. In the first place, the supposed decline of human semen quality is still controversial. The paper of Carlsen *et al.* has received much criticism and other papers trying to discern a time-trend in semen quality show contradictory results. In the second place, there is no evidence for a causal relationship between environmental estrogens and human semen quality. Wildlife effects cannot be readily extrapolated to the human situation because there is a major difference between exposures of wildlife and man. In the third place, the link between sperm quality and the actual male fertility is unclear and there is little evidence that male fecundity is declining (Sherins 1995).

The purpose of the present study was to critically analyze the data presently available on time trends in sperm quality, to assess the likelihood of a true decline, and to discuss the implications of the findings for government policy.

2. LITERATURE SEARCH PROCEDURE

2.1 Publications of interest

The purpose of this study was to evaluate the likelihood of a change in sperm quantity or quality by means of a literature search and analysis. An inventory of factors that may influence sperm parameters and sperm measurement methods was made beforehand to establish criteria for the analysis (chapter 3). Information on the factors that influence sperm parameters and on sperm parameter measurements were found in the literature. Additional practical information on sperm measurement methods was obtained orally from Dr R.F.A. Weber, Erasmus University Hospital, Rotterdam.
The meta-analysis by Carlsen *et al.* (1992) was discussed as this study has a central place in the ongoing discussion on the putative change in sperm quality (chapter 4). In addition, publications were selected and analyzed in which a time trend over the years was analyzed by the same study center (chapter 5).

2.2 Search and selection procedure

Publications on time-related measurements of sperm quantity and/or quality were identified as follows: MedLine (SilverPlatter) was searched with key words 'decreas* semen quality', 'decreas* sperm quality', 'declin* sperm count*', 'human sperm count*', 'human male fertility', 'chang* semen', 'chang* sperm quality' and 'chang* male fertility'.
Recent articles were found using Current Contents On Diskette (CCOD) with key words 'sperm', 'semen', 'fertility'. All publications that had a title that could indicate usefulness for the present study were looked up, the relevant articles were selected. Reference lists of the found articles and editorials, letters and reviews (for example Toppari *et al.* 1995) were searched to make sure all relevant publications would be found.
One recent article was obtained with the help of Dr. Weber. The most recent article on the subject could be included in this report with the help of Dr. C.E.J. Cuijpers, RIVM.

2.3 Results of search and selection

18 Publications were found that described a trend in sperm parameters over time, observed in one lab with an approximately similar population. Many different countries were represented: the United States, Denmark, Sweden, South-Africa, Norway, France, Belgium, United Kingdom, Finland, Greece and Italy.
Publications that observed a decrease in sperm quality or quantity were Leto & Frensilli (1981), Bostofte, Serup & Rebbe (1983), Osser, Liedholm & Ranstam (1984), Menkveld *et al.* (1986), Bendvold (1989), Bendvold *et al.* (1991), Auger *et al.* (1995), van Waeleghem *et al.* (1996), Irvine *et al.* (1996), Adamopoulos *et al.* (1996) and Menchini-Fabris *et al.* (1996)
Publications that did not observe any decline were MacLeod & Wang (1979), Wittmaack & Shapiro (1992), Bujan *et al.* (1996), Paulsen, Berman & Wang (1996), Vierula *et al.* (1996), Fisch *et al.* (1996) and Fisch *et al.* (1997).

Letters of Irvine (1994), Ginsberg *et al.* (1994) and De Mouzon *et al.* (1996) presented data on a changing semen quality as well, but were not analyzed because these letters did not supply enough information.

3. CRITERIA FOR EVALUATION OF SCIENTIFIC PUBLICATIONS ON SPERM PARAMETER TRENDS

The results described in a publication are valid only if the epidemiological study has been designed and performed properly. There are two aspects regarding the results obtained:
1. Precision, dependent on within- and between-subject variation, measurement inaccuracies and the accuracy of the outcome measure. This is dependent on the number of individuals or samples included in the analysis.
2. Validity. An epidemiologically incorrect design could cause wrong conclusions through confounding and bias. These alter the relation that is under investigation, in this case, the relation between sperm parameters and time.
Even if hypothesis testing reveals significant results, both validity and precision could provide alternative explanations for the observed findings (Hennekens & Buring 1987). Only when results for a certain population are found to be valid, the question whether results can be generalized to the whole population can be considered.

In order to be able to evaluate the suggestions in literature that sperm quality or quantity has decreased during the last decades, the validity of the results of the meta-analysis of Carlsen *et al.* (1992) and the 18 other publications was analyzed. Each publication was reviewed with regard to epidemiological design, possible confounders, sperm measurement methods and statistics to see whether the conclusions are legitimate. The approach used is described at the end of this chapter. But first, some information is provided on sperm measurements, possible sources of bias and confounding, and statistics.

3.1 Sperm parameter analysis

3.1.1 Sperm parameters

Quantitative parameters, the most commonly measured parameters in the publications, are sperm concentration (= sperm density, millions/ml), semen volume (ml) and the composite of the two, total sperm count per ejaculate (millions). Qualitative parameters include measurements on sperm motility, morphology and viability (% live) (Bonde *et al.* 1996). For these, several different parameters are used. Morphology can be described by %normal sperm, %abnormal sperm, or more detailed: %normal heads, %amorphous heads, %double heads, %pathological midpeaces etcetera (Sheriff 1995). Motility is commonly graded *a* to *d*, from grade *a* representing sperm with rapid progressive motility to grade *d* representing sperm with no motility (WHO 1992). %Motility is defined as the percentage of sperm that shows any motility (grades *a*, *b*, and *c*). %Progressive motility is the percentage of sperm with grades *a* and *b* motility, %rapid progressive motility is the percentage of sperm with grade a motility. Although these are the most commonly used measurements others have been used as well, depending on the authors (Leto & Frensilli 1981, Irvine *et al.* 1996)

3.1.2 Variation in sperm parameters

Sperm parameters vary within and between individuals. For example, a within-subject variability of 55 to 75% of the total variability of motility measurements was found in a population with known medical and occupational histories (corrected for age and

duration of abstinence) (Schrader *et al.* 1991). However, for all motility parameters except %motility a single sample was found to be representative of both population and individual values. Mallidis (1991) observed that 50 to 75% of all variability in sperm concentration, semen volume and motility index was within-subject variability. This intraindividual variability can be regarded as a result of random measurement error around the true individual mean and temporary influences such as infections, stress, seasonal variability and abstinence (Mallidis 1991, Tielemans *et al.* 1997). Between subject variation can be very high as well for some sperm parameters, for instance sperm concentration (Sherins 1995). This variation seems to be less in a population of fertile men and higher in infertile men and the whole population (Sherins 1977). Here again, the variability has several causes: biological differences, differences in life-style, duration of abstinence, age etcetera.
To reduce the inaccuracies in measured sperm parameters that result from both within- and between-subject variability, a study should include a sufficient number of men. This number depends on the specific study population and on the study design. Taking multiple samples per man reduces within subject variability as well (Berman 1996). Every publication will be evaluated as to whether the number of men and samples is sufficient, without doing specific power-calculations.

3.1.3 Sample collection

Samples should be collected by masturbation. Coitus interruptus is no good collection method because part of the ejaculate can be missed (McLeod & Wang 1979; Sheriff 1995) and because psychological factors have a large influence on resulting sperm quality (Le Lannou & Griveau 1996). Sperm motility in a sample decreases with time, therefore analysis of the sample must be performed as soon as possible, preferably within two hours after ejaculation (Sheriff 1995). All samples need evaluation within this time interval otherwise differences in the outcome will result.

3.1.4 Sperm analysis methodology

Lack of standardized laboratory procedures, subjectivity and inaccuracies of the various methods and investigators bias the results of sperm analyses (James 1980; Seracchioli 1995). To reduce the differences, the WHO recommended unified techniques for semen analysis in their manual published in 1980 and revised in 1987 and 1992. But even these WHO guidelines could not prevent sperm measurements to differ within and between labs. Small changes in the methods of measuring sperm parameters, such as changes in the use of pipette tips, can bias the outcome of routine semen analyses (Knuth *et al.* 1989 in: van Roijen *et al.* 1995). Moreover, even the WHO manual does not give an unequivocal description of what constitutes a 'normal' sperm morphology. Normality has probably been assessed with increasingly strict criteria over the years in many laboratories as a result of increasing focus on changing sperm parameters and the introduction of an alternative morphology-scoring method: 'strict criteria' (Kruger *et al.* 1986). An indication of this can be found in Tielemans *et al.* (1997). They observed a highly reduced %normal sperm in men attending an infertility clinic in 1991 compared with 1974-1984 although it was expected to have increased due to higher fertility of the men in 1991 (see paragraph 3.2). As a result of the probable change in the assessment of normal morphology, publications that observe a decreased %normality should be regarded with care.
In each of the publications addressing a possible time-trend, sperm analysis is either performed directly upon delivery of the sample or on cryopreserved samples. The

results of analyses directly upon delivery are subject to the disturbing influence of changes in materials and methods over time such as the ones already mentioned. This makes the comparison over time difficult. To be trustworthy, a publication should describe the methods employed and assure the reader that validation (internal quality control) took place to guarantee precise and consistent measurements (Mortimer *et al.* 1986). Cryopreservation can possibly alter sperm parameters. Therefore cryopreserved samples cannot be compared with fresh samples. If such a comparison is made in a publication, the results cannot be trusted.

3.1.5 Measurement inaccuracies

Matson (1995) found coefficients of variation of 5% to 66% in the measurements of sperm concentration in four replicates in one lab, with average values around 15%. Neuwinger *et al.* (1990) observed similar results and noted that the variation is highest in samples with low concentrations and lowest in samples with high concentrations. They further found that especially morphology (%abnormal) measurements and single motility categories (like *a*, *b* and *c*, in contrast with overall motility and progressive motility) are less precise. Motility and normality are to some degree subjective parameters. Only %immotile and %normal heads are relatively objective parameters and can be determined more precisely.

The results found by Matson (1995) for sperm concentration differed somewhat depending on the counting chamber used: the highest value (66%) was found using a Makler chamber, the lowest (5%) with a Neubauer chamber. Mortimer *et al.* (1986) observed a variation of less than 5% when performing duplicate analyses with an improved Neubauer heamocytometer. They mention that for other counting chambers which do not require dilution of the aliquot, such as the Makler chamber and the Horwell chamber, coefficients of variation ranging from 21.6% (low concentration range) and 6.1% (high concentration range) were observed by the designers. However, these measurements were not performed according to the same procedures. Not enough evidence was found for considering one counting chamber to be superior to another. However, it is clear that results obtained in one counting chamber should not be compared with results obtained in another type of chamber because this can lead to results differing 30% (van Roijen 1995).

Neuwinger *et al.* (1990) observed a variation of 10% between technicians for sperm concentration. This was found to be largely due to sampling error because variation decreased to 2 % when the heamocytometer was filled once and counted 10 times instead of filling it 10 times. Similar very small variations between technicians were found by Mortimer *et al.* (1986). Jequier & Ukombe (1983) observed a much higher variation between (experienced) technicians: 37.8% for sperm concentration. However, these (three) technicians from different labs were asked to perform the analysis in the manner with which they were most familiar, so part of the variation could be due to different methods. Therefore we expect no important influence of the employment of different technicians as long as they are well-trained.

Coefficients of variation of 23 to 87 % are found between labs evaluating sperm concentration, morphology and motility (Neuwinger *et al.* 1990). Until good external quality control is established, no relevant comparisons between the results of different laboratories can be made. Therefore, if in a publication data obtained in different labs are compared over time, conclusions cannot be trusted because of possible observer bias. Some have stated that comparisons between methods or labs should not be made

based on coefficients of variation, but on the methods described by Bland and Altman (1986) (van Roijen *et al.* 1995).

3.2 Bias

Bias may be defined here as any systematic error in an epidemiological study that results in an incorrect association between sperm quality and time. Bias can occur when the methods of recruitment or selection of participants at different times are 'different' and this difference is related to sperm quality, or when non-response (motivation to participate) changes over time and is dependent on fertility. Population characteristics that influence sperm quality are for example fertility status (fertile, infertile, fertility unknown (Sheriff 1995, Comhaire *et al.* 1996)) and geography (general term for genetic/ethnic, environment and life style influences). In the case of a birth-cohort study bias occurs when the older population differs from the younger population with regard to fertility. One more possibility for obtaining wrong results is observer bias. Observer bias is introduced when measurements obtained by different observers are compared, as could be noted in paragraph 3.1.4.

An example of selection bias: Several databases exist on men with infertile marriage because their semen was analyzed in the course of an infertility investigation. The population of men of infertile couples that have their semen analyzed has almost certainly changed over time (Bendvold 1989, Bendvold *et al.* 1991). In early years (sixties, seventies) men of infertile couples were only asked to have their semen analyzed after thorough investigation of the female partner. Nowadays men are asked to provide a semen sample at the beginning of the infertility investigation as a routine procedure. It can therefore be expected that 50% of the men of infertile couples nowadays are of 'normal' fertility whereas in early years less of the checked men were of 'normal' fertility. The actual quantitative change in semen parameters that occurred over time is difficult to estimate and depends on the practice of each different clinic. Infertility can be caused by a whole array of factors that exert their influence in different ways: the infertility could be based on any single semen parameter that was subnormal, with all other parameters being normal. Sample sizes have to be sufficient for randomizing for this heterogeneity.

3.3 Confounders

Confounding involves the risk that the observed association is due, totally or in part, to the effects of differences between the study groups other than the determinant under investigation. In the case of sperm measurements over time, these are factors that influence sperm parameter characteristics. When these factors are not recorded and if necessary corrected for, it is possible that an observed association between sperm parameter values and time is due to the effects of a change over time in the values of the confounder(s). What factors are possible confounders and how they influence sperm parameters is described below.

3.3.1 Age

There seems to be general agreement on the negative effect of increasing age on sperm motility (Schwartz 1983; Wang *et al.* 1985; Haidl, Jung & Schill 1996). Some found a decrease in %motility with every earlier year of birth, others found values decreasing in

men aged over 46-50 and under 25 (Schwartz 1983). Publications are less unanimous on the influence of age on the other sperm parameters. Schwartz (1983) found a reduced percentage of normal cells with men aged over 41 and under 25. Lower sperm concentrations were found by Haidl *et al.* (1996) but not by Schwartz.
Because the study of Schwartz is a qualitatively good study, it will be assumed here that there are good indications that sperm motility and normality increase until the age of 25, remain high until the age of 45-50 and subsequently decline. Consequently, confounding is expected only in studies in which the proportion of old or young men has changed during the period under investigation. When age is not taken into account at all, confounding cannot be ascertained but the results cannot be trusted to describe a real trend either.
Age can be a confounder for duration of abstinence as well because longer abstinence periods could possibly be found more commonly in older men. This is only important when duration of abstinence is not taken into account.

3.3.2 Duration of abstinence (ejaculation frequency)

The longer the duration of sexual abstinence, the higher sperm concentration, volume, total sperm count, and the lower % motility (Pellestor *et al.* 1994; Wang *et al.* 1985) and % normality (Pellestor *et al.* 1994; Mortimer 1985). Values of the quantitative parameters can differ threefold between 1 and 7 days of abstinence. The percentages motility and normality were not observed to decline after 2 or 4 days of abstinence, only after 7 days a significant reduction was observed (Pellestor *et al.* 1994). Magnus *et al.* (1991) also found that the percentage motility only decreased after abstinence of 7 to 10 days. So, duration of abstinence influences all sperm parameters, with quantitative parameters changing earlier and more pronounced than the qualitative ones.
The ejaculation frequency influences values in sperm analysis as well. Frequent ejaculation (every one or two days) halves the ejaculation volume and causes a lower sperm concentration, with full recovery only after 4 days or more of abstinence (Tyler *et al.* 1982; Nnatu *et al.* 1991). Oldereid *et al.* (1992) also found a decrease in the percentage abnormal sperm and an increase in the percentage progressive motility when increasing the ejaculation frequency from <1 to >4 times a week. In contrast, Tyler *et al.* and Nnatu *et al.* did not find any effect on motility or morphology. The resulting confounding effect of ejaculation frequency is estimated here to be small however: when a man with a high ejaculation frequency is asked to refrain from sexual activity for 3 days, the values for the sperm parameters at the beginning of the abstinence period will be relatively low but the recovery rate (daily sperm output) is higher than in men with a low ejaculation frequency, who begin the abstinence period with higher values for the sperm parameters (Le Lannou & Griveau 1996).

Taken together, it is evident that the results of a study can have been subject to major confounding if durations of abstinence were not recorded or changed over time. The number of days of abstinence before sampling that was requested by the investigators has no important meaning because actual durations are often distinct from the asked durations and asked durations are not precise, for example 3 to 5 days.
In case abstinence was not recorded, it is interesting to know in what way confounding could have influenced the results. Bendvold *et al.* (1991) observed a decrease in duration of abstinence in infertile couples in Norway from 7.5 days in 1956 and 1966 to 5.0 days in 1976 and 4.4 days in 1986. Westoff (1974, in James 1980) as well found

an age dependent increase in marital coital rate in the United States during 1965-70. James (1996) found an increase in coital rates of 22% between 1965 and 1975 in married couples in the US and a subsequent fall between 1975 and 1988 of 27%. Trends in the general population are less well known. James (1996) proposed that recently a reduction of coital rate has occurred as a result of the growing perception of the danger of aids. However, no evidence was presented. It is therefore clear that duration of abstinence underwent substantial changes over time and that publications which do not take abstinence into account could be subject to major confounding regardless of the time period studied.

3.3.3 Season of sampling

In several publications sperm concentration values were observed to oscillate in a year's time, with highest values generally found in winter and lowest values in summer (Tjoa *et al.* 1982; Levine 1991; Schrader *et al.* 1991). Vierula *et al.* (1996) found this variation for volume and total count as well. Mallidis *et al.* (1991) in contrast did not find any changes. The subject therefore is not quite clear yet but all publications together seem to indicate that changes occur over the seasons, at least in temperate zones. As long as samples used in a study are randomly obtained over the year, no confounding influence is expected. Only when sampling season differs between groups, such confounding is possible.

3.3.4 Other possible confounders

Several studies found a detrimental effect of *smoking* on sperm concentration and morphology (Chia *et al.* 1994; Vine *et al.* 1994). This was found in both fertile and infertile men. Smoking seems to reduce sperm concentration with approximately 15%. Other studies only found a reduced ejaculation volume with smokers (Osser *et al.* 1992; Holzki *et al.* 1991) Furthermore, there are several publications that report no influence of smoking (review in Oldereid *et al.* 1992). Consequently, the existence of a detrimental effect of smoking on sperm quality is still controversial.
Excessive *alcohol use* does influence spermatogenesis (Oldereid *et al.* 1992, Bonde *et al.* 1996). Presumably, a threshold of 50g alcohol/day has to be exceeded before any toxic effects on the testes can be manifested (Oldereid *et al.* 1992).
Radiation and *stress* were observed to have a negative influence as well (Bonde *et al.* 1996).
A whole series of articles exist on the possible influence of *profession* (see Tas 1996). Certain professions (welders, agricultural workers, professions with exposure to radiation but also chemicals like ethylenebromide, glycolethers and lead) are associated with reduced sperm quality and quantity (Bonde *et al.* 1996). A higher frequency of poor semen parameters was recently reported both in white collar workers and in transportation workers such as drivers of taxis and buses (which spend a lot of time sitting) (Figà-Talamanca *et al.* 1992). These sitting professions probably exert their influence through the slight increase in scrotal temperature which impairs spermatogenesis (Bonde 1996b). It should be considered that most findings were derived from a small number of cases.

Varicocele is regarded by some as the main reason for infertility (Weber 1995). Prevalence in the standard population of healthy men is estimated to be 15%, up to 40% in men attending infertility clinics. Varicocele influences percentage normal morphology, total and progressive motility and vitality (Paduch & Niedzielski 1996).

Other factors known for their negative effect on sperm are *fever*, *sexually transmitted disease* (Purvis & Christiansen 1993 in: de Waal *et al.* 1995), genital abnormalities like *hypospadias* and *cryptorchidism* (Yavetz *et al.* 1992 in: de Waal *et al.* 1995) and *infections*. In more general terms, varicocele, hypospadia and fever have their detrimental effects through the negative influence of elevated temperature on sperm quality.

If these possible confounders such as illnesses or life-style are not mentioned for the population under study, it is difficult to judge whether confounding has taken place. Confounding is more probable when only two moments are compared (for example 1970 and 1990) than when measurements were performed continuously (from 1970 to 1990), because the chance of having two distinct populations with regard to that and other unknown confounding factors is very probable. This is important especially when only a small number of individuals is included.
On the other hand, if selection was deliberately carried out to prevent confounding, the representativeness of the selected population for the general population is questionable.

3.4 Statistics

For quantitative sperm parameters the distribution of individual measurements is skewed towards lower values (Bromwich *et al.* 1994; Keiding 1994). The median therefore is more accurate than the mean for describing the data. With this non-normality of data, it is probable that the variance is dependent on the mean. Therefore the data should be transformed for example with a log-transformation when a regression analysis is performed (Berman *et al.* 1996). For correlation analysis transformation is only necessary when using the Pearson correlation test.
Some sperm parameters are classified, like progressive motility. This parameter rated from 0 to 4+. This is an ordinal scale, meaning that the distances between the classes are not equal. Medium (2) for example, is not twice as good as bad (1). As a consequence, one cannot calculate mean and variations on these variables, nor compare these incorrectly calculated means for different populations by using a student t-test. Only non-parametric methods can be used (Sheriff 1995).
The degree of variance declared is an important measure when regression analysis is performed. When the degree of variance declared is very small, the regression line accounts for only a small proportion of the variance in the data (that always show large variation). The probability of this line representing a true (biological) decline in such case is low.

3.5 Procedure for analysis of publications

The results of the analysis of the meta-study by Carlsen *et al.* (1992) is described in chapter IV. This analysis regarded possible bias and confounders and the sperm measurement methods and statistics. Some additional criteria specific for the analysis of a meta-study are described at the beginning of the chapter concerned.
Chapter V contains the analyses of all 18 publications on a time trend in sperm quality or quantity. Each article is discussed as follows:
First, the contents of an article with regard to

- study design

- study population
- sample size
- age characteristics of the study subjects
- abstinence characteristics of the study subjects
- other possible confounders that are taken into account (illnesses and life-style)
- used sperm measurement methods
- results, with the used statistics

is shortly described in a table. Then comments are provided in which possible sources of bias and confounding are regarded, as well as the consistency in sperm measurement methods, epidemiological design and statistical methods used. Finally, an overall conclusion on the validity of the study with respect to its outcome regarding a trend in sperm quality is formulated.

4. THE META-ANALYSIS BY CARLSEN *et al.*

In 1992 Carlsen and colleagues published a meta-analysis in which they concluded that human sperm quality decreased significantly between 1940 and 1990: mean sperm concentration from $113*10^6$/ml to $66*10^6$/ml and seminal volume from 3.4 ml to 2.75 ml (Carlsen *et al.* 1992). In this chapter the results of a close examination of the methods of the meta-study are presented, to see whether the conclusion is legitimate. It is important to note here that the authors use the expression 'sperm quality' for sperm concentration and ejaculate volume, whereas 'sperm quantity' would be better. Hereafter 'quantity' instead of 'quality' will be used to describe these parameters.

4.1 Methodology of a meta-analysis

In a meta-analysis, publications about a specific subject are compared in order to reveal new or more detailed information. Not all available papers can be used for such a purpose. To guarantee the meta-analysis describing a real value or trend instead of an artifact, only comparable papers can be used. This means ideally that:
1. Studies should deal with identical populations.
2. Studies should use the same measuring methods
3. Confounding factors in the studies should be comparable
However, it is also possible to use papers that are not fully comparable by adjusting for these differing factors, so that they become comparable.

A good protocol for conducting a meta-analysis should include the following steps (de Hollander *et al.* 1996):
a. Definition of the objective of the study.
b. Definition of the search procedure to be used.
c. Definition of in- and exclusion criteria. In this way, papers that meet the constraints mentioned above can be selected.
d. Select the qualitatively good studies. ('If you put junk in, only junk will come out')
e. Make a conscious choice about the statistical methods that will be used.
f. Check for major influences of individual papers on the results.

Focusing on the subject of sperm quantity, studies included in a meta-analysis should be comparable or at least adjustable for the following factors:
1. Population:
    - fertility (unknown fertility, fertile, primary and secondary infertile, father)
    - geography (and ethnicity)
2. Determination of sperm quantity parameters.
3. Confounding factors:
    - age
    - duration of abstinence and ejaculation frequency
    - life-style (smoking, drinking, profession etc.)
    - illnesses (fever, infections, varicocele)
    - season of sampling

4.2 The study of Carlsen and colleagues

4.2.1 Objective and search procedure

The objective of the study is stated as 'to investigate whether semen quality has changed during the past 50 years'. To identify appropriate papers, the authors searched in *Medline Silver Platter* database for studies published during 1966 to august 1991, with the key words: sperm count, sperm density, sperm concentration, male fertility, and semen analysis. For studies published between 1930 and 1965 they *used Cumulated Index Medicus* (or *Current List* 1957-9, covering the three years when the index was not published) with key words spermatozoa, semen, and fertility. This search procedure must have yielded several hundreds of potentially useful publications.

4.2.2 In- and exclusion criteria

Only studies of humans were selected. No further inclusion criteria were used. Of the hundreds of potentially useful publications the search procedure produced, only some are about sperm quantity. A quick search of *Cumulated Index Medicus* showed that there are some articles that seem useful (for example 'Determination of reproductive capacity in male by examination of sperm' (Nippe, Deutsche Ztschr.f.d.ges.gerichtl.Med. 26:64-69 '36) and 'Observations on specimens of human semen' (Tynen, J Contraception 4:125-127 may '39)) although this cannot be deduced from the title alone. How the selection of useful articles was made is unclear and it is unknown whether all relevant articles were included in the analysis.

Publications were excluded if
a. they included men from infertile couples or those referred for oligospermia or some genital abnormality,
b. they included men selected for either a high or a low sperm count, and
c. counting of sperm cells had been performed with a computer assisted system or flow cytometry.
The articles were included irrespective of sample size, as a result of which there are 2 publications with less than 10 men, 21 publications with less than 30 men, and 32 publications with 50 men or less (more than half of the publications!). The distribution of sperm concentration measurements of individuals is highly skewed towards the lower values. In such cases the use of medians is preferable to the use of means (Bromwich *et al.* 1994; Keiding 1994). In the meta-analysis means are used because not all publications mentioned median values. But then again, not all publications mentioned means either, so three papers were included by their median values (numbers 16, 57 and 60 in Carlsen's table). Instead, these papers should have been excluded.

It is apparent that no effort has been made to select publications that are comparable in terms of population, sperm analysis method or confounding factors. Also, there does not seem to have been a selection of qualitatively good studies. The selection procedure led to 61 articles with a total of 14,947 men. From these articles the authors recorded the following data: year of publication, country of origin, and information about the men with respect to possible fertility, age, and (when available) race. Furthermore they recorded mean, median and ranges of sperm densities (with standard deviations) and seminal volume and period of sexual abstinence.

4.2.3 The dataset

Let us look now at the characteristics of these 61 publications and see whether populations, measuring methods and confounding factors are comparable or corrected for. To do this, we decided not to review all 61 papers because this would take too much time. Instead we took all papers with 1000 or more individuals included, since Carlsen and colleagues weighted the publications by number of subjects in their analysis. These papers are: McLeod & Gold (1951), Naghma-E-Rehan *et al.* (1975), Tjoa *et al.* (1982), Sheriff (1983), Wang *et al.* (1985) and Pol *et al.* (1989).

4.2.3.1 Population

Fertility class and subpopulation: the meta-analysis included data on men unselected with respect to fertility as well as data on fertile men. Selection bias can very well have occurred here since only the 6 biggest studies already include several different subpopulations. The group classified as fertile (Carlsen *et al.*) consists of:
1. Men reporting for vasectomy having 2 to 4 children, and the last conception less than 6 months ago (Sheriff 1983),
2. Men reporting for vasectomy having 2 or more children and some of them having their wife pregnant (Rehan *et al.* 1975),
3. Volunteers for artificial insemination having at least one child (Pol *et al.* 1989),
4. Men with a pregnant wife (McLeod & Gold 1951).

The group of men unselected with respect to fertility consists of:
1. Sperm donors + pre-marital checkups + men reporting for vasectomy. 1157 of these men were of unknown fertility, 82 had children (Wang *et al.* 1985),
2. Men reporting for vasectomy, of which 95% was fertile (criterion not mentioned) (Tjoa *et al.* 1982).

It can be expected that fathers, fertile men, men with unknown fertility (definitions in Sheriff 1995) as well as sperm donors differ with respect to their sperm parameters (Sheriff 1995; McLeod & Wang 1979; Sherins *et al.* 1977). Moreover, knowledge of fertility problems has increased with the years so nowadays more women can become pregnant in comparison with the past, thanks to treatment of these problems in both women and men. This means that the population of fathers has changed.
Carlsen *et al.* state that separate analysis of the papers with only fertile men led to the same conclusions, so joining the groups did not alter the conclusions. Apart from the fact that this does not justify the inclusion of so many different subpopulations, one can wonder why they did not analyze the group of men unselected with regard to fertility. This group is even more varied than the group of fertile men. Is it possible that they performed the analysis indeed but found that it did not fit expectations?
Bromwich *et al.* (1994) as well as Olsen (1994) put forward that a changing definition of normal sperm concentration during the period under examination can cause selection bias. Formerly, sperm concentration was considered abnormal if lower than $60*10^6$/ml, whereas nowadays the cut-off value is $20*10^6$/ml. This can have caused exclusion of some men in the early publications. This fact alone, they say, could account for the found reduction in sperm quantity. Carlsen *et al.* oppose this view by pointing out that early publications do in fact contain lots of below-normal values (Keiding 1994a), which seems to be true.

Geography and ethnicity: publications originating from all over the world are included, with almost half originating from the United States. This setup has the disadvantage

that measurements of sperm quantity carried out in different areas and at different times, are compared. No good conclusions can be drawn from such a comparison: it is impossible to eliminate the possibility that the difference found has its origin in a geographically varying sperm quantity. This difference can have several causes: it can be due to ethnical (genetic), life-style or environmental factors. In the paper of Carlsen and colleagues information on race was supposed to be noted, but they found that most publications did not provide this information.
Fisch & Goluboff (1996) reanalyzed Carlsen's data to see whether geographical variations influence the findings. They used only those papers from the meta-analysis which included 100 men or more. 20 publications remained, containing still 91% of all men in the meta-analysis. They found that all 5 papers before 1970 (these are the ones with high values for sperm quantity) originated from the United States, with 4 of them from New York. After 1970 (when the lower values are found), only 3 out of 15 papers originated from the United States and only one was from New York. Moreover, of the 7 papers with the lowest values found, 5 originated from developing countries (not represented in the early period).
The results of McLeod & Wang (1979) confirm the possible influence of geography. They reviewed publications on sperm concentration of fertile men in the United States for the period 1938-1977. Although these populations too were not fully comparable, there are some interesting findings: the results of the three early studies (the same as used by Carlsen) are comparable and high. They all contain New York data. The four later studies are not comparable in spite of describing the same period: some contained high values, others contained low ones. But we find these papers originating from different areas: Iowa, Houston and New York (2). One of the New York papers and the Houston paper do not figure in the analysis by Carlsen. What we found, is that the New York results are indeed comparable. Also, the Houston-publication (Smith & Steinberger 1977) measured approximately the same sperm concentration as another Houston-publication used by Carlsen *et al.* (Tjoa *et al.* 1982): $70*10^6$/ml and $66*10^6$/ml. A reanalysis by Bahadur *et al.* (1996) led to the conclusion that the decline in sperm counts is largely accounted for by these USA data and not the European and Asia/Africa/South American data.
In conclusion we can say that there are important indications that there are geographical differences in sperm concentration. The trend in sperm quantity found by Carlsen *et al.* could entirely be accounted for by these geographical differences.

4.2.3.2 Determination of sperm quantity parameters:
The outcome of sperm parameter measurements is very sensitive to changes in methods (van Roijen 1995, Matson 1995, Sheriff 1995). In the past there were no standardized laboratory procedures available for the analysis of sperm. Only in 1980 the WHO presented the first manual on this subject, aiming to make sperm measurements reliable and comparable between labs and countries. It can therefore be expected that considerable differences in methods and materials exist between recent and old studies (Ulstein 1996). One possible assumption is that recent techniques are more accurate, which would lead to a higher value of sperm quantity parameters measured. When in spite of this a decrease is found, one can wonder if the decline in reality would not be bigger still. However, there are so many methodological factors influencing the outcome of an analysis that can differ between studies, that no comparisons between these outcomes can be made. These methodological differences render the analysis of Carlsen *et al.* unreliable.

4.2.3.3 Confounding factors:

Age: information on age was provided in only 42 of the publications. When looking at 6 of the papers, the age-interval (number of ages in the study) varies from 13 to 42 years: from 28 to 40 years old in Sheriff (1983) and from 23 to 64 years old in Rehan *et al.* (1975). The other papers too differ as to what extent ages associated with reduced sperm quantity (under 25 and over 45 years old) are included. In conclusion, the populations are not comparable with regard to age, neither has there been a correction for the influence.

Duration of abstinence and ejaculation frequency: Only 2 out of 6 papers noted the exact durations of abstinence: in one paper this ranged from 1 to 60 days (with a mean of 6,5), in the other it ranged from 3 to 30 days. A third paper mentioned an abstinence of 3 days, but it is not clear weather this is a prescribed or a true period. In the 3 remaining papers a period of 3 to 5 days, 2 to 5 days, or 2 days was recommended, but the actual period was not recorded. No paper contained any information on ejaculation frequency. Since sperm concentration depends heavily on both duration of abstinence and ejaculation frequency, the value of a comparison of these studies is questionable.

Life-style, illnesses and other confounding factors: 4 out of the 6 papers in Carlsen *et al.* regarded here do not provide information on these subjects. Rehan *et al.* (1975) recorded ethnicity as well as religion and sexual habits. In the study of Wang *et al.* (1985) men who had chronic mental illness, a history of abnormal puberal development, cryptorchidism, drug or alcohol abuse, exposure to toxic chemicals or irradiation were excluded. Thus, we can presume that different studies used different inclusion and exclusion criteria with regard to confounding factors. Because only some of the studies mentioned confounding factors, they cannot be chequed on comparability. There is a chance that especially the smaller studies are an unrepresentative sample of the underlying population, with different values for confounding factors.

4.2.4 Statistical methods

After having regarded the weaknesses of the publications included in the meta-analysis, also a lot can be said about the statistics that are used. The publications are not evenly distributed over time: for the period 1930-1970 only 13 publications are included, against 48 papers for the period 1970-1990. Of these 13 early papers only 5 contain more than 50 individuals. Taking into account the rarity of early publications it is disputable whether statistical methods can lead to valid conclusions on a trend for the whole period. This critique is also expressed by Olsen *et al.* (1995) and Brake & Krause (1992).

Even when we assume that there are enough data on the period until 1970, the validity of the conclusions is questionable. Carlsen *et al.* used linear regression to describe the relation between sperm count and year of publication (and also between ejaculate volume and year of publication, but since no additional information on this relation is provided in the article, it is not regarded here). The use of this model has received much criticism in the literature. Some authors put forward that a biological parameter can rarely be described by a linear regression. More often an exponential, logarithmic or cyclic models is useful (Farrow 1994). Olsen *et al.* (1995) reanalyzed the data using a set of different statistical models and found that all models (for example the stairstep-

model) had a better fit than the linear model. These models all led to the conclusion that from the sixties on there had been no further reduction or even a slight increase in sperm count. Several other authors noted that the apparent reduction seemed to be step-wise (before and after 1960) instead of linear (Brake & Krause 1992; Joffe 1996). In response to the criticism Carlsen and her co-authors admitted having found equal results, without mentioning the reason for choosing the linear model for their publication (Keiding 1994).
Other authors that reanalyzed the Carlsen-data (Bahadur *et al.* 1996) found that the linear regression line was no longer useful for describing the data when the most recent reports (Auger *et al.* 1995, Irvine *et al.* 1996 and Bujan *et al.* 1996) were included. The quadratic model had a better fit and suggested that there was a gradual rise in sperm count since 1975. However, Bahadur *et al.* did not include all recent studies in their re-analysis, such as Wittmaack 1992 and van Waeleghem 1996. It is unclear why, since they do mention these studies in their discussion.
The linear regression is based on year of publication instead of year of measurement. Carlsen and colleagues noted that the year of measurement was up to 10 years before the year of publication in some papers, but still used this as the time parameter. This is a major confounding factor. For the six largest papers duration of study varied from 3 to 7 years and publication took place between 0 and 5 years after measurements had been terminated. What's more, in two papers (Tjoa *et al.* 1982; Pol *et al.* 1989) we found time spans between measurement and publication up to 11 years.

Several alternative explanations for the described difference in sperm concentration values before and after 1960 exist. Not only the changed methods of sperm measurements but also changed sexual practice after the sexual revolution in the sixties (including the introduction of birth-control pills) could explain the differences. It is probable that the sexual revolution led to shorter periods of abstinence (Olsen 1994; Joffe 1996). With shorter abstinence intervals lower sperm concentration values would be found. Carlsen *et al.* could not find any support for this view, but Westoff (1974, in James 1980) reported an age-dependent increase in marital coital rate in the United States during 1965-1970. Bendvold *et al.* (1991) too found that the sexual abstinence in their research objects declined from 7.5 days in 1956 and 1966 to 5 days in 1976 and 4.4 days in 1986. Another possibility (though without any evidence) is a natural fluctuation throughout the years. To be able to support this, information on the period before 1938 would have to be available.

Carlsen *et al.* identified the 10 publications with the greatest influence on the estimate of the regression coefficient and concluded that none of them differed substantially from the other papers. These 10 papers turned out to be the 9 largest papers in the meta-analysis (evidently a consequence of the fact that papers were weighed by number of subjects included) and the paper with the highest sperm count. Although the difference between these and the other studies cannot be judged here, the value of the 6 biggest studies was found to be unconvincing. Leaving out the largest paper (Tjoa *et al.* 1982) containing 29.6% of the total number of subjects included did not have any effect on linear regression line (Bahadur *et al.* 1996).

The concentration ranges in Carslen's report ($<20*10^6$/ml, $20-40*10^6$/ml, $41-60*10^6$/ml, $61-100*10^6$/ml and $>100*10^6$/ml) were regarded to detect frequency changes over time. These are important parameters regarding male fertility: values

below $20*10^6$/ml are considered abnormal by the WHO (WHO 1992). The conclusion of Carlsen and colleagues was that high concentrations had reduced in frequency, whereas the low concentrations had become more common. However, only 27 of the 61 publications are used for this analysis, without any specification on which ones these are. Representativeness therefore cannot be judged. Figure 2 in Carlsen *et al.* shows that the conclusion is based on information about 439 individuals. The most puzzling is that for the period 1951-1960 89 men are mentioned in the figure, but the only two available publications in this period contain 21 and 1000 individuals. The value of this analysis is questionable.

4.3 Conclusion

Summarizing, the paper by Carlsen and colleagues does not provide convincing evidence for the conclusion that sperm quantity has decreased over the past 50 years. There are several reasons for this conclusion:

1. The relative scarcity of publications before 1970.
2. Large variation in the dataset due to differing population fertility, life-style factors, duration of abstinence, ejaculation frequency, and the use of year of publication of the data instead of year of measurement.
3. Confounding with geographical differences, changed sexual habits and differing sperm analysis methodologies.

The observed trends in the sperm concentration ranges and ejaculate volume could not be judged as a consequence of inadequate or insufficient information.

5. ANALYSIS

5.1 Publications that observed a decrease in sperm quality or quantity

5.1.1 Leto & Frensilli (1981).

Changing parameters of donor semen. *Fertil Steril*, 36, 766-70.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of data obtained between 1973 and 1980 (one year intervals) in the US.	Potential sperm donors that came to the study center from nearby medical, law, dental and graduate schools. 80% students. Selection criteria unaltered. The population was split in two: accepted and rejected donors. To be accepted, donors had to have good sperm (minimal count $75*10^6$/ml, at least 50% motility etc.) and be free of veneral diseases.	150 accepted donors (14 to 27 per year), 135 rejected donors (7 to 36 per year). Accepted: mean of 12 monthly samples used. Rejected: mean of 3 monthly samples.	-	3 days asked

Other confounders	Sperm analysis	Results
Accepted donors free of veneral diseases.	One technician. A summary of the techniques was given. Equipment was periodically calibrated.	Student t-test used. **Sperm density** showed a gradual decline for both accepted and rejected donors. No change in **%motility**. **Forward progression** declined only for accepted donors and only between 1978 and 1980. **%viability** became significantly different from 1973 in 1979, but only for the accepted donors. **%normal** forms decreased for the accepted donors and increased for the rejected donors. All significances $P<0.05$. More potential donors had to be rejected each year on the basis of the criteria for acceptance.

Commentary:
Methods of recruitment were not specified. It is possible that these or (non-)response changed over time.
Men were classified as donor only if all sperm parameters were according to the criteria. Therefore rejection could have been based on either of all parameters, leading to enormous heterogeneity in this group. The rejected donors delivered 3 samples. It is not clear which or how many were used to decide rejection.
Confounding with age is possible. 80% of the men were students and therefore probably aged 18 to 25. No information was provided on the other 20% of the men or on a possible change in mean age during the study period.
Confounding with duration of abstinence is also possible because actual values were not recorded. These sources of confounding can not only have influenced the observed time-trend, but also have led to misclassification of rejected and selected donors. The fact that in the group of accepted donors more significant relations were observed is possibly because this group is more homogenous than the other. The value of the declines observed still cannot be judged because of possible confounding.
Because for the accepted donors a mean of 12 samples per man was used, within-men variability was largely reduced. But the groups of accepted and rejected donors did

both include a very small number of men in almost each year, as a result of which the between-men variability could have influenced the outcome measure to a great extent. No time 'till analysis was mentioned. If this time was not strictly defined this could have influenced the results for the motility measurements. The decline in %motility was observed using a parametric test regardless of the fact that this is a non-parametric measure. Therefore the significance of this small change in forward progression is questionable. What's more, all observed trends were significant at the $P<0.05$ level, but means instead of medians were used so the real declines could as well be non-significant.

Conclusion:
Response and selection bias possibly occurred. Confounding with age and duration of abstinence could have influenced classifications and therefore the trends observed. Even the fact that more potential donors had to be rejected each year could be due to confounding. The use of means instead of medians makes the significance of the observed trends (all $P<0.05$) doubtful. Therefore the results can certainly not be trusted to describe a real decline in sperm quantity and quality.

5.1.2 Bostofte, Serup & Rebbe (1983).

Has the fertility of Danish men declined through the years in terms of semen quality? A comparison of semen qualities between 1952 and 1972. *Int J Fertil*, 28,91-95.

Study design	Population	Sample size	Age	Abstinence
Retrospective comparison of data obtained in 1952 and 1972 in Copenhagen.	Men examined because of a fertility problem, from Copenhagen and suburbs. In '52 12.7% was from the suburbs, in '72 4.2%.	1077 in 1952 and 961 in 1972.	Men in 1972 were 3 years younger (medians), with more men aged under 20.	-

Other confounders	Sperm analysis	Results
In 1972 more high-class men were included.	Performed by one technician or under his strict supervision. Methods, according to R. Hammen's methodology, remained unchanged. Bürker-Türk chamber.	Non-parametric methods used. A non-significant decrease in **volume** and no change in **%immobile** spermatozoa was observed. **Sperm density** (median 73.4 to 54.5*10^6/ml, P<0.01) and **motility** decreased and **%abnormal** sperm increased (26% to 44%, P<0.01).

Commentary:
Selection bias towards higher sperm quality/quantity is expected because more 'normal' men are expected to be included in the population in later years.
Duration of abstinence was not recorded. Shorter periods are expected in 1972.
Confounding with age is also possible for %immobile and %abnormal sperm.
The influence of differences in social status between the groups is unclear. This regards occupational as well as lifestyle factors.
With only two times of measurement that are 20 years apart ('52 and '72), small changes in methodology could have occurred unnoticed and have great influence.
It is remarkable that %immobile sperm did not change but overall motility was found to have reduced. In general, the classification of motility is much more subjective than the assessment of %immobile sperm and therefore conclusions on this parameter should be regarded with reservation.
Why did the authors take 1972 to compare with 1952 while they performed the analysis/published the study in 1983. It seems strange to wait 11 years before publishing the article. Is it possible the authors picked two years that suggested a decline when compared instead of two other years that would not suggest such a decline?

Conclusion:
It is unclear why more recent data were not included in the study. Confounding with abstinence and selection bias have likely occurred. Possible confounding with age, changes in methodology and discrepancies in %motility versus %immotility further suggest that the results obtained cannot be trusted to describe reality.

5.1.3 Osser, Liedholm & Randstam (1984).

Depressed semen quality: a study over two decades. *Arch Androl*, 12, 113-116.

Study design	Population	Sample size	Age	Abstinence
Retrospective comparison of sperm obtained in 1960-61 and 1980-81 in Sweden.	Men from infertile couples, living in the city of Malmö or surrounding areas.	158 men in both 1960-61 and 1980-81.	21-64, mean 31. Subjects in both groups were age-matched.	3 to 5 days asked.

Other conf.	Sperm analysis	Results
-	Smears of '60-'61 were reassessed. Bürker chamber. One technician.	Multiple regression analysis was used because of patterns of internal correlations in the material. Significant reduction of **volume** (median decreased from 3.8 to 3.4 ml) and **sperm density** (mean from 109 to 65 $*10^6$/ml) and increase in **%double heads** observed. The overall **%abnormal** sperm did not change. Separate analysis of samples from men from Malmö and from the surroundings showed no significant **volume** change for either group. Decline in **sperm density** seemed larger in the city. Surroundings: still significant increase in **%double heads** found, but decreases in total **%abnormal** sperm and **%amorphous** heads.

Commentary:
In both years samples were provided at the start of the investigation of presenting infertile couples. Duration of infertility was at least one year for all couples. Therefore no selection bias is expected.
Confounding with duration of abstinence can explain the difference found between 1960 and 1980: the abstinence period probably changed to a lower value. The differences that were found between values in Malmö and in the surroundings could as well be due to confounding with abstinence.
The statistical methods remain vague: it was mentioned that multiple regression analysis was used because of patterns of internal correlations in the material, but what variables were used, how this was done or what internal correlations were concerned is not described in the article.

Conclusion:
Confounding with abstinence could explain the difference found between the two years and between the city and its surroundings. This together with the uncertainties regarding statistical methods renders the validity of the conclusions questionable.

5.1.4 Menkveld *et al.* (1986).

Possible changes in male fertility over a 15-year period. *Arch Androl*, 17, 143-144.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of data obtained over the period 1968-1982 in South-Africa.	Men attending an infertility unit, 50.8% whites, 42.1% coloureds, 6.4% blacks and 0.7% Asians. And some donors.	4459 primary or secondary infertile men and 322 donors.	-	-

Other confounders	Sperm analysis	Results
-	-	Averages of all semen parameters (**volume**, **concentration**, **%motile** sperm, **speed of forward progression** and **%normal** sperm) were calculated separately for each group for each year, Asians were excluded due to small numbers. Correlations for the semen parameters and years 1968-82 calculated: notable correlation with time only found for **%normal** morphology: whites -0.34, coloureds -0.42, blacks -0.39 and donors -0.57.

Commentary:
The study design has serious flaws as no notice was taken of confounders.
There is a possibility of selection bias: it is mentioned that from 5007 men attending the unit, 4459 were included in the study. Men were said to be primary or secondary infertile, but no definitions were given. Normally, infertility in the absence of fatherhood is considered primary, whereas infertility occurring after fathering a child is considered secondary. It is not clear whether a selection was performed on the men at the infertility unit with regard to the fertility of the wife, because no inclusion/selection criteria were mentioned.
The ethnicity of donors is unknown (or at least not mentioned).

Conclusion:
The results of this study are highly doubtful because no confounders were taken into account and no methods were described. Furthermore, selection bias could have occurred.

5.1.5 Bendvold (1989).

Semen quality in Norwegian men over a 20-year period. Int J Fertil, 34(6), 401-404.

Study design	Population	Sample size	Age	Abstinence
Retrospective comparison of semen parameters in 1966 (actually '66-'69) and 1986 in Norway.	Men from infertile couples providing sperm for routine analysis, almost exclusively from Oslo and surroundings.	1966: 125 first samples 1986: 129 first samples, men were randomly selected.	Distribution was similar for the two years.	3 to 4 days recommended.

Other confounders	Sperm analysis	Results
-	All morphological smears restained and reevaluated by one technician (criteria by Bjøro and Strand). Volume and concentration (Bürker chamber) were measured by the same leading doctor and technician.	All values were classified, Wilcoxon signed rank test used. Significant change over time in **%normal** morphology (from 60% in 1966 to 41% in 1986, $P<0.001$). No changes observed in **volume** and **sperm density**.

Commentary:
Because a population of men from infertile couples was investigated, selection bias is expected towards higher sperm parameter values in 1986. Men were randomly selected, but out of what 'pool' is not described. This as well can have caused selection bias.
The measurements being performed on only two moments in time, together with the relatively small sample size and the heterogeneity of the study population, could have interfered with the validity of the findings.
Duration of abstinence was not recorded.
No sample collection method was mentioned.
Volume and concentration measurements were performed by two individuals. Even if one assumes that both analyzed a similar proportion of the samples in each year, it cannot be excluded that slight changes in the analysis of samples have occurred between 1966 and 1986.
For morphology measurements at least methods were similar because for these all smears were reassessed.

Conclusion:
Because of possible confounding with duration of abstinence and possibly occurred methodological changes and selection bias, the validity of the findings in this article is questionable. Moreover, measurements were performed in only two years and sample size was relatively small taking into account the heterogeneity of the group of men from infertile couples. This could have introduced large differences between the populations in the two years that influenced sperm parameter measurements.

5.1.6 Bendvold, Gottlieb, Bygdeman and Eneroth (1991).

Depressed semen quality in Swedish men from barren couples: a study over three decades. *Arch Androl*, 26(3), 189-194.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of data measured in 1956, 1966, 1976-79 (autumn) and 1986 in Sweden.	Men from barren couples providing a samples as a routine procedure in investigation. Most from Stockholm and surroundings.	141, 201, 219 and 224 men respectively (first samples).	1956: 33.4 (21-56) 1966: 31.0 (18-55) 1976: 32.1 (20-57) 1986: 34.5 (22-61)	Reduced from 7.5 to 4.4 days during study period. For analysis men were grouped: abstinence of 3-5 days or 6 days or more (only for '56 and '86).

Other confounders	Sperm analysis	Results
-	Bürker chamber. Morphology assessed according to Eliasson and WHO '80 criteria.	Wilcoxon signed rank test used. **Total sperm count** reduced significantly between 1956 ($460*10^6$) and 1986 ($305*10^6$) in the whole group. But when abstinence was taken into account no decline was found for men with 6 days or longer abstinence. For the men with 3-5 days abstinence a significant (P=0.05) decline was found. **%Normal** decreased stepwise between 1956/66 and 1976/86. Sexual abstinence did not influence the tendency in %normal.

Commentary:
In both 1956 and 1986 the men were divided in groups with either 3 to 5 days of abstinence or 6 or more days of abstinence. The number of men included in each of the different groups was not mentioned. The change in abstinence duration in addition to the considerable
variation in abstinence within subgroups affects the credibility of data over time. The conclusion that %normal was not influenced by duration of abstinence was based on the same abstinence-classification mentioned above and is therefore subject to doubt. Confounding with age is possible for the %normal sperm but would be very small. The difference in age characteristics of the population over the years is not that large.
No information was provided on how many technicians were employed or on their training. Regarding the length of the study period, several technicians must have been employed. Therefore, differences between the periods can have been introduced. The use of different criteria (Eliasson and WHO) could have introduced substantial differences.

Conclusion:
The results obtained in this study cannot be trusted because of methodological changes and confounding with abstinence. In addition, no information was given on the employment of technician(s).

5.1.7 Auger, Kunstmann, Czyglik & Jouannet (1995).

Decline in semen quality among fertile men in Paris during the past 20 years. *N Engl J Med*, 332, 281-285.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of data obtained between 1973 and 1992 in Paris.	Healthy fertile sperm donors, 95% white and 85% living in the Paris area. All fathers.	1351 men.	Mean age at donation increased from 32 in 1973 to 36 in 1992.	3 to 5 days asked, adjusted for in analysis.

Other confounders	Sperm analysis	Results
The group of 1750 men out of which the 1351 study objects were taken were manual workers, technicians, executives.	Analysis within 1 hr. During the study time 11 technicians worked at the lab. They had the same training. Unchanged methodology which was regularly verified, no new equipment introduced. Methods for measuring volume, concentration and %normal described, motility was assessed according to WHO '80.	Linear regression: sperm concentration decreased from $89*10^6$/ml in 1973 to $60*10^6$/ml in 1992. Multiple regression analysis: older age contributed significantly to the decreases in concentration, %normal and %motile sperm. Abstinence increased with increasing age. Adjustment for age and abstinence revealed a decline in **concentration** of 2.6% and of 0.3% and 0.7% in **%motile** and **%normal** sperm, respectively, associated with each successive year of birth. **Volume** did not change.

Commentary:
Apart from donors being healthy and fathers, no selection criteria were mentioned, which could conceal selection bias.
The professions within the group of 1750 men (the included donors, vasectomy patients and brothers of infertile couples delivering a sample for artificial insemination) was given, but their representation in the actual 1351 men included in the analysis was not mentioned. Therefore, the information given on profession is not useful.
11 Similarly trained technicians worked in the lab, which does not seem a source of bias.
The morphology assessment was according to WHO guidelines, but these did not exist up to 1980 so a change is very plausible. Therefore the claim of the authors, namely an unchanged and verified methodology for the assessment of all parameters, seems to be questionable.
The scatterplots of sperm density and %normal morphology are very unconvincing to support a decline. The proportion of variance declared should have been mentioned.

Conclusion:
Not only methodological changes but also selection bias may have occurred. The results of this study therefore cannot be ascertained to describe reality.

5.1.8 Irvine *et al.* (1996).

Evidence of deteriorating semen quality in the United Kingdom: birth cohort study in 577 men in Scotland over 11 years. *Br Med J*, 312, 467-471.

Study design	Population	Sample size	Age	Abstinence
Retrospective birth cohort study on data collected between 1984 and 1995 in Scotland.	Scottish volunteer donors for the unit's program of gamete biology research. Recruitment at antenatal clinics, with advertisements to undergraduate populations and by personal contact via existing donors.	577 first ejaculates This is more than 100 per birth cohort. Cohorts: '55-9, '60-4, '65-9 and '70-4.	No evidence found that age of donors had changed during the period of data collection with respect to the year of first donation.	3 to 4 days asked.

Other conf.	Sperm analysis	Results
No difference in smoking or drinking habits between the cohorts.	In general, analysis was according to WHO guidelines, within 90 minutes. Methods described. Improved Neubauer chamber.	Few of the younger birth cohorts were of proved fertility. Linear regression: significant decrease in **concentration** (2.1%/year), **total number** of sperm (2.01%/y), **overall motility** (0.18%/y) and **total number of motile** sperm (2.04%/y) with later year of birth. No significant change in **volume**. Similar relations but with opposite sign were observed between age at donation and semen quality, except for overall motility (no change). Stepwise multiple regression with both age at donation and year of birth showed that mostly only year of birth was negatively related to semen quality (only total number of motile sperm seemed to be more strongly, positively related to age). Conclusion: semen quality is deteriorating.

Commentary:
Selection bias has almost certainly occurred and could explain the observed results. The men recruited at antenatal clinics were probably older and with proven fertility, whereas the students were younger and mostly of unknown fertility.
No information was given on employment and training of the technicians. The effect of the difference between technicians would have been somewhat reduced because the data were analyzed by birth-cohorts, causing samples analyzed by different technicians to end up in various cohorts..
As not all analyses were performed according to WHO guidelines, discrepancies may have occurred.
On the basis of absence of a change in ejaculate volume with age, the authors concluded that duration of abstinence did not differ between the birth cohorts. This conclusion is questionable in view of the possible disturbing influence of year of birth and changes in methods of sperm measurement. Duration of abstinence could therefore very well be a confounder.

Conclusion:
Selection bias may have caused the observed decline in sperm quality with later year of birth. This means that a steady sperm quality in reality would be a just as plausible conclusion based on these results, which cannot be ascertained because of possible confounding with duration of abstinence and possible methodological bias.

5.1.9 Van Waeleghem *et al.* (1996).

Deterioration of sperm quality in young healthy Belgian men. *Human Reprod*, 11, 325-329.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of data obtained between 1977 and 1995 in Belgium.	Candidate sperm donors recruited through advertising in local newspapers and student periodicals. The majority unmarried university students or paramedic personnel who had not fathered a child.	416 first samples. Accepted donors delivered a 2nd sample.	20-40, 90% being between 20 and 30. Did not change over the study period.	3 to 5 days asked.

Other conf.	Sperm analysis	Results
Candidates were asked to present only when in good health, without history of serious illness and with a negative history of illness in their family.	WHO '87 criteria used, for conc. measurements the method according to Hellinga (1976) as count estimation was used as well. Nearly all analyses performed by the same technician, using the same method and registration form. Analysis < 1hr.	Correlation and ANOVA: no significant (P=0.08) decrease in **concentration** with time, ANOVA: P=0.12. Significant increase in **volume** (P<0.01, ANOVA P=0.045). Dual dot plots comparing '77-'80 with '90-'95: significant decrease of **concentration** (P=0.035), slightly increased **volume** (P=0.067). **Total count** did not change. **%normal** morphology (R=0.046, coeff. of determination), **%rapid motility** (R=0.185) and **%progressive motility** decreased, **%immobile** increased (all P<0.0001). Rapid motility seems to have stabilized recently because of a higher R for the quadratic fit. A significant linear correlation found between the results of sperm motility and morphology. The second samples provided by the accepted donors were compared with the results of the first samples and showed no sign. difference for concentration and morphology. Total motility was slightly better (P=0.02).

Commentary:
Samples were collected the whole period through, except for a gap of almost two years between 1988 and 1990. Why this occurred is not described in the article. This coincided with the introduction of the WHO-manual (1987) which may have resulted in a change in sample measurements. In the scatter diagrams of sperm concentration and %normal morphology in relation with time, a discontinuity in obtained values before and after the gap is visible. Only the scatter diagram of %rapid progressive motility in relation with time seems to indicate a decrease over the whole study period. %rapid motility and %immobile were not represented in figures, therefore they cannot be evaluated. The coefficients of determination are very low and were not even mentioned for all parameters. So, the obtained results do not really support the conclusion of a deteriorating sperm quality.
The division of men into two groups ('77-'80 and '90-'95) for analysis was not substantiated.
Sample size was approximately 21 samples (men) a year, allowing a large influence of within- and between-subject variation. The authors themselves found that for the total population of accepted donors over all years (which is much more uniform than the population of candidate donors) the second samples significantly (P=0.02) differed from the first samples for total sperm motility. For total count, seminal volume,

progressive and rapid motility the differences between first and second samples were not mentioned.
Observer bias may have been introduced by the participating of a second technician in sample analysis.

Conclusion:
As a result of methodological changes, the analysis of samples in two arbitrary groups and the dubious representativeness of the values obtained for several parameters, the results of this article cannot be trusted to describe reality.

5.1.10 Menchini-Fabris *et al.* (1996).

Declining sperm counts in Italy during the past 20 years. Andrologia, 28, 304.

Study design	Population	Sample size	Age	Abstinence
retrospective comparison of data obtained in Pisa, Italy between 1975 and 1994.	men chosen among those consulting at the department of andrology to check their fertility.	4518 men. Difference in number of samples was minimized.	differences in age were minimized. Mean age decreased from 30.8 in 1970 to 29.9 in 1990.	differences in abstinence were minimized.

Other confounders	Sperm analysis	Results
only men presenting with no andrological disease affecting fertility after objective examination were selected.	technicians did not change during study, one lab probably. No methods mentioned.	Results were analyzed in three groups: from 1975 to 1979, 1983 to 1986 and 1991 to 1994. Ejaculate **volume** showed a slight decrease (from 3.2 to 2.9ml), **sperm count** declined from 71.8 to $65.32*10^6$/ml. **Forward motility** fell from 50.1% to 32.1%.

Commentary:
The characteristics of the study population were not defined and may have changed with time.
It was mentioned that 'differences in age, abstinence and number of samples were minimized according to Sherins (1995) and Bromwich *et al.* (1994)'. On the basis of this information it is impossible to assess whether all possible confounding with age and abstinence was ruled out.
The ejaculate volume varied to such an extent over the three periods (3.25, 2.47 and 2.96 ml respectively) that confounding and/or bias of any kind seems very likely.
Sperm count and forward motility only declined between the first and the second period, they remained stable afterwards.
No information was given on the use of medians and on transformations applied.
Methodology could not be evaluated because it was not mentioned.

Conclusion:
This study was presented in very short format and failed to address a number of crucial aspects necessary to evaluate the validity of the results. The likelihood of confounding and selection bias together with a possibly changed methodology and the use of means instead of medians for sperm count render the results of this study doubtful.

5.1.11 Adamapoulos *et al.* (1996)

Seminal volume and total sperm number trends in men attending subfertility clinics in the Greater Athens area during the period 1977-1993. *Hum Reprod*, 11(6), 1936-1941.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of data obtained between 1977 and 1993 in Greece, grouped per year.	Male partners of subfertile couples in the Greater Athens area, of Greek origin.	2385 men randomly chosen out of 23850 men (one out of each 10).	22-53. No marked differences in mean age among years.	3 to 5 days asked.

Other conf.	Sperm analysis	Results
-	Analysis in three labs with similar methods, each lab one technician. Coefficients of variation calculated and found small. Data on motility were not included in the analysis since they were not uniformly documented at the initial stage of study and therefore were inconsistent with the criteria set by WHO ('80) at least for a number of years.	No distinct relationship found between age and seminal **volume** either in individual years or for the whole sample and the total number of years. Parameters analyzed using linear and non-linear (cubic, quadratic) regression. A declining tendency in volume (cubic,R^2=0.34, P<0.05) found. The **percentage of men with sub- or supranormal volume** was relatively stable (using WHO-criteria). A marked decrease in mean **total sperm number** noted from year 1986 onwards and a decline over the total period (cubic,R^2=0.67, P<0.001). **%Men with subnormal total sperm number** (<40*10^6/ml) did not change significantly. Only after pooling of groups (<120 and >240*10^6/ml) changes became significant. Analysis of data from individual laboratories showed no significant trend in seminal volume but marked differences for total sperm count.

Commentary:
Confounding with duration of abstinence is possible since actual values were not recorded.
Samples were analyzed in three laboratories. Because all laboratories performed analyses during the whole study period in the same region with the same methods, only one technician performed the analyses in each laboratory and the inter-observer variation was small, no observer bias is expected.
Selection bias is probable towards higher sperm quantity in later years because a population of men with subfertile marriage was investigated.
Small changes in sperm measurement methodology could have occurred because no validation took place.
In the early years three outliers in total sperm count were found: 1650, 2700 and 2950*10^6. Nothing was mentioned on the possibility that these were produced by error and no separate analysis without these samples was performed to evaluate their weight.
The best-matching curves for volume and total sperm count are consistent with a fluctuation of the parameters over the years.

Conclusion:
Selection bias towards higher values would not invalidate the observed decline in sperm quantity, but confounding with abstinence and possible changes in sperm measurement methodology shed doubt on the usefulness of the findings.

5.2 Publications that did not observe any decline

5.2.1 McLeod & Wang (1979)

Male fertility potential in terms of semen quality: a review of the past, a study of the present. *Fertil Steril*, 31, 103-116.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of data obtained between 1966 and 1977 in the US, grouped in 9 periods. These data are compared with data obtained in 1951.	Men with infertile marriage and men referred from elsewhere because their semen quality was analyzed and found wanting (=second examination).	9000 (9 groups of 1000 subsequently examined men), 5476 second examination (first 500 patients seen each year).	-	Great majority of specimens obtained after 3 days abstinence.

Other confounders	Sperm analysis	Results
-	Prior to 1972 all counts were obtained by technicians using standard hemocytometer techniques (in '51 as well). Subsequently, counts have been done by the senior author.	Ejaculate **volume** was constant over the years. Difference in median sperm **concentration** values between 1951 and 1966-1977 was very small and non-significant. **Count frequency distributions** were similar as well. The 9 periods measured between 1966 and 1977 (infertile marriage) differ significantly (X^2, $P<0.005$) in their count frequency distributions (higher values in period 7 and 8, lower in period 5). However, a persistent trend was not present for the means, medians or frequency distribution. The sperm concentrations of patients referred because of poor semen quality were significantly lower than of the men with infertile marriage. Similar elevations (in period 7 and 8) were found.

Commentary:
Observer bias is possible because of the replacement of the former technicians by a new one. It is not certain that these different technicians performed the counts in exactly the same way. It is mentioned that standard hemocytometer techniques were used. This is a very broad term and a changed methodology can therefore not be excluded.
Selection bias might have occurred, but since the values for the population of men that had their second examination vary in the same manner (high values in period 7 and 8, low in period 5) as the infertile marriage-population, this is not probable.
In the absence of exact data on abstinence period prior to sampling, confounding with abstinence may have influenced the results of the study.
Confounding with age is not expected because of no known effect of age on sperm concentration and volume.

Conclusion:
Observer bias and changed methods, together with confounding with abstinence shed serious doubt on the usefulness of the data in this study.

5.2.2 Wittmaack & Shapiro (1992)

Longitudinal study of semen quality in Wisconsin men over one decade. *Wisconsin Med J*, 91, 91-95.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of data obtained between 1978 and 1987 in Wisconsin, randomly distributed throughout the year.	Potential sperm donors. They answered advertisements placed on the university campus. They had to have a college degree, be in good health, but did not have to be married or of proven fertility.	159 men, 4 to 26 per year.	-	Samples were collected after 72 hours of abstinence.

Other confounders	Sperm analysis	Results
-	Analysis within 30 min.	Student t-test used. Ejaculate **volume** did not change. No statistical trend in **concentration** or **%motile** sperm found. %abnormal sperm rose sharply between 1982 and 1983. This coincided with a change in criteria used to identify abnormal sperm. In the pool of accepted donors a similar sudden change was found.

Commentary:
A sample size of 4 to 26 men/samples per year is too small to eliminate disturbing influence of the large intra- and interindividual variations and introduces extensive deviations from the actual means. The trustworthiness of a trend over the years derived from these obtained values is low.
No specification of the term 'in good health' was given. Variable interpretation of this term may have introduced response bias, but this seems unlikely because a change over time should have occurred to introduce bias.
Samples were supposed to be collected after 72 hours of abstinence. However, it is not clear whether this is an asked or a recorded time so confounding is possible.
No information was given on the age characteristics, which may have influenced %motility. Confounding is therefore possible.
Nothing was mentioned on the methods of sperm analysis or possible changes in employment or training of the technicians. Changes therefore could have occurred but cannot be evaluated.

Conclusion:
This publication has little value for assessing the likelihood of a trend in sperm quality because of an inadequate sample size. In addition, possible confounding with duration of abstinence and age and unspecified sperm analysis methods and employment of technicians render the results unreliable.

5.2.3 Bujan, Mansat, Pontonnier, Mieusset (1996)

Time series analysis of sperm concentration in fertile men in Toulouse, France between 1977 and 1992. *Br Med J*, 312, 471-2.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of samples obtained between 1977 and 1992, grouped per year. In France.	Healthy unpaid candidate sperm donors at the Toulouse clinic. All had previously fathered at least one child. Donors aged less than 20 and over 45 were excluded, as were donors with an infertile brother.	302 first samples. 11 to 27 per year.	Mean age at donation increases over the time of the study	3 to 5 days asked.

Other confounders	Sperm analysis	Results
Healthy donors. Place of residence mentioned.	Methods described in Mieusset *et al.* (1987).	Linear regression between **sperm concentration** and year of donation showed a significant positive correlation ($P<0.05$). However, when adjustment was made for the donor's age, the relationship was no longer significant. Multiple regression analysis including year of birth and age at donation revealed that only the donor's age contributed significantly to sperm concentration: 3.3% increase for each year increase in age. Conclusion: no change in sperm concentration in Toulouse.

Commentary:
The term 'healthy donors' was not specified. No recruitment methods were mentioned either. Selection bias is therefore possible.
The small sample size per year sheds serious doubt on the validity of the yearly averages and therefore on the trend over the years.
No information was given on the employment and training of the technicians. Analysis was on defrosted samples indicating standardized methods for all samples.
Duration of abstinence was not taken into account. It is very well possible that the observed increase in sperm concentration with increasing age is a result of confounding with abstinence, for older men can be expected to have longer durations of abstinence and a longer duration of abstinence is commonly related to a higher sperm concentration. Therefore the adjustment for age may actually have been an adjustment for duration of abstinence, which seems even more plausible because no effect of age on sperm concentration was expected on the basis of existing literature.

Conclusion:
The small sample size sheds serious doubt on the validity of the findings in this article. Moreover, selection bias and observer bias could have occurred.

5.2.4 Paulsen, Berman & Wang (1996)

Data from men in Greater Seattle area reveals no downward trend in semen quality: further evidence that deterioration of semen quality is not geographically uniform. Fertil Steril, 65(5), 1015-1020.

Study design	Population	Sample size	Age	Abstinence
retrospective time-analysis of data obtained between 1972 and 1993 in Greater Seattle area	normal, healthy men donating semen samples as participants in 14 clinical studies. Nearly all white, most graduate students or recent graduates.	510 men. Median of 5 to 7 samples p/man used	18-52, 95% between 18-40. No change over time.	2 to 7 days asked.

Other confounders	Methods of analysis	Results
no heavy tabacco use, alcoholism or drug abuse reported. No chronic systemic illness, obvious varicoceles, hormonal deviations. No history of infertility. No agricultural or manufacturing jobs.	one lab, no change in personnel. Morphology assessment according to WHO '80. Volume: volumetric pipette. Conc: Coulter counter, with previously validated method. All methods were consistent.	Regression analysis: **Sperm concentration, volume, total count** and **%normal morphology** did increase significantly. However, the degree of variance declared was very low (1 to 7% explained by time change). Conclusion: no time trend in the Greater Seattle area.

Commentary:
Selection bias is possible. Recruitment methods and inclusion criteria were said to be similar for all 14 studies. However no information is provided on the similarity of the resulting populations. For instance, some men were recruited by means of their wives who were attending a prenatal clinic. It is not known how these fertile men are distributed over time.
The proportion of men under 25 was slightly higher in the early years. This could have influenced %normal sperm towards higher values in later years, but since the difference between the years is small, no important influence is expected.
Duration of abstinence was not recorded and the asked interval was very broad. The use of multiple samples per man helped to partly reduce variability caused by different abstinence periods. This does not preclude confounding completely, for mean durations of abstinence differ between men as well. The possible direction of this confounder is unknown, as there are no clues on how duration of abstinence varied between 1972 and 1993.
Since an unchanged method and validation were mentioned, methods of sperm analysis probably remained unchanged between 1972 and 1993.

Conclusion:
The results of this study could be subject to confounding with abstinence and selection bias. The validity of the findings can therefore not be ascertained.

5.2.5 Vierula *et al.* (1996)

High and unchanged sperm counts of Finnish men. *Int J Androl*, 19(1), 11-17.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of data obtained between 1967 and 1994 in Finland.	Infertile men providing semen samples as a routine procedure in the investigation of infertile couples. Most were from the city of Turku and it's surroundings.	5253 first samples. Mean/year =196, range 46-298.	Increased from 28 to 33 during the study period.	3 to 5 days asked, adjusted for in analysis.

Other confounders	Sperm analysis	Results
No influence of life style expected.	WHO '87 methodology used. Both Bürker and Bürker-Türk chambers used. Neither methods nor technicians changed during the study period.	Multiple regression analysis revealed, after adjustment of duration of abstinence, age and season of sample delivery to the year of sample delivery, a decreased semen **volume** (4.5 to 4.1 ml, $P<0.001$) and no significant changes in sperm **concentration** or **total sperm count**. Similar results were found when adjustment was made to year of men's birth instead.

Commentary:
Selection bias towards higher values is expected because of the use of a population of men from infertile couples. Sample size is large enough to reduce the influence of the heterogeneity of the group.
WHO methodology was supposed to be used, but the study includes results of semen analyses performed between 1967 and 1994. Although the authors wrote that methods did not change, small changes could very well have occurred. In addition, different counting chambers were used which is expected to have introduced extra variation. Because of the long time span of the study, it would be expected that more than one technician would have performed the analyses.

Conclusion:
Selection bias and observer bias as well as methodological changes could have influenced the results obtained in this article.

5.2.6 Fisch *et al.* (1996)

Semen analyses in 1,283 men from the United States over a 25-year period: no decline in semen quality. *Fertil Steril*, 65(5), 1009-1014.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of data obtained between 1970 and 1995 in the US.	All men banking sperm before vasectomy in Minnesota ('70-'94), New York ('72-'94) and California ('78-'94). Minnesota and New York: 64% proven fertility (=had fathered a child), 24% unproven fertility, 12% did not answer the question. In the California data fertility status was not available.	1283 men: Minnesota: 662, New York: 400, California: 221. Some delivered more samples (in such cases mean values used).	Minnesota men were of lower age (33.0) than New York (35.6) or California (36.2) men. Mean age at specimen collection increased over the study period from 30 to 38.	Significantly lower in California. It did not change over the study period or with increasing age.

Other conf.	Sperm analysis	Results
-	Minnesota: Neubauer chamber, California: Makler chamber, New York: first Neubauer (72 to 77) then Makler chamber. Motility assessment methods described for Minnesota and California. Analysis within 1 hr.	The three sperm banks differed in age, abstinence and mean semen characteristics and were included in the study at different times. In a multiple regression analysis was controlled for these factors by correcting for the changing proportion of subjects reported from each site from year to year with the differing means of semen parameters. Still a (small) increase (P=0.004) in **concentration** was found, but no change in **motility** or **volume**. Linear regression analysis showed a significant increase in sperm concentration in Minnesota and New York, but not in California. Linear regression analysis by fertility status (only Minnesota and New York) showed a significant increase in concentration for the 'proven' and 'unproven fertility' groups, no significant change was observed in the group with unknown fertility.

Commentary:
No selection bias is expected from the combination of data from men with differing fertility since fertility status did not influence inclusion in the study.
Characteristics of the population with regard to fertility may have changed over the years.
Age at specimen collection increased significantly over the study period. This can be an artifact of the fact that the Minnesota-studies were included from 1970 on, while the Californian studies were included only after 1978. No information is given on a possible age-trend in the separate regions.
Confounding with abstinence could only have occurred if this changed differently in the separate regions. This is unlikely because no trend was found in the overall data.
The differences between the labs in counting chambers and technicians employed were corrected for in the multiple regression. An influence of change of technicians within one or more labs was however not excluded.
Although for the individual regions it was not investigated whether mean age or abstinence changed over time, the linear regressions seem to indicate an increase in

two of three regions. The fact that no significant change was found for the men of unknown fertility may be related to the heterogeneity of this group. Sample size in California was too small to find any trend in this region alone.

Conclusion:
Data from three distinct areas were combined without checking for the possibility that geographical differences exist in trends in sperm parameters, methodology, abstinence or age. Moreover, the characteristics of the population might have changed over the years. This could have altered the relation under investigation.

5.2.7 Fisch *et al.* (1997)

The relationship of sperm counts to birth rates: a population based study. *J Urol*, 157, 840-843.

Study design	Population	Sample size	Age	Abstinence
Retrospective time-analysis of data obtained between 1971 and 1994 in Minnesota, US.	All men banking sperm before vasectomy during this time. 57% fertile, 29% no children, 14% fertility unknown.	660 men, with a total of 1972 specimens. Analysis was on specimen-level.	Mean age increased from 29.7 to 36.5 (recorded for each specimen).	Mean value was 4.4 days, no change over study period (recorded for each specimen).

Other confounders	Sperm analysis	Results
Monthly variation of parameters taken into account.	Sperm counting technique was according to WHO and remained unchanged throughout the study period.	Multiple regression analysis adjusted for age, duration of abstinence and season (month) revealed a significant linear increase in the mean annual **sperm concentration** ($1.03*10^6$/ml year, $P<0.0001$). Significant effects of abstinence and month, no effect of age observed. Controlling for the linear trend, a significant, non-linear year-to-year fluctuation was found. Concentration was found to be highest in march and lowest in September. When mean sperm counts of the 660 men were used in analysis, similar results were found.

Commentary:
The study population might have changed over time because of changed motivations or selection criteria.
No validation of the sperm measurement methodology took place. Small but important changes could therefore still have occurred.
No information was provided on how many technicians were employed and how they were trained.
It is possible that men provided different numbers of samples and therefore were weighed differently in analysis. This would only have biased the results if it concerned specific men with high or low fertility, which is not likely.

Conclusion:
No information was provided on the technicians and selection and/or methodological bias may have occurred. Therefore the validity of the results cannot be ascertained.

6. GENRAL DISCUSSION

6.1 Reporting bias

During the last decades, many studies have reported on mostly negative time-trends in sperm parameters. The meta-analysis by Carlsen *et al.* (1992) was a hallmark in this list. Not only did it receive relatively much attention in the media, it also initiated the publication of several reports in which the absence of a trend or an increased sperm quantity was reported.
The rarity of studies reporting no trend before the appearance of the article by Carlsen *et al.* may be the result of reporting bias. Before the fuss that was created by the article of Carlsen, there was no interest for such reports. In this period only two out of eight articles published on the subject reported an unchanged sperm quantity/quality (McLeod 1979 and Wittmaack 1992). After the enormous publicity that was given to the subject, studies that observed an unchanged sperm quantity/quality suddenly became very meaningful and were published. The fact that this resulted in half of the articles reporting no change or an increase and half of them reporting a decline, might indicate that sperm quality/quantity has not changed, even with these studies not being performed very well.

6.2 Methodological issues

Analysis of the separate publications showed that, although all authors presented their results as good evidence for either a decline or no change, many of them did not regard various methodologic aspects that may invalidate their conclusions. The differences between the results of the separate publications could be caused by these factors.
It was notable that in most publications sperm analysis methods did not receive much attention. Either it was described summarily or not at all and there were cases in which data obtained with different methods were compared. Internal quality control, or validation of the methodology, was only mentioned in some articles. The importance of strict standardization of sperm parameter assessment was not recognized in the majority of the studies. The possibility that changes were introduced by the subsequent employment of several technicians was not recognized either as many articles did not even mention the subject.
In most publications recruitment methods or selection criteria were not described, which could have introduced selection bias if these changed over time. Some publications claimed to have found a trend in sperm parameters although sample size was that small that between-subject variation would have caused considerable deviations of the obtained values from the actual means in the population.
In many studies confounders such as age and abstinence were not taken into account, especially in the earlier ones. As described in chapter 3, several indications were found in the literature on possible changes in abstinence period during the last decades. However, these were not completely consistent and concerned a specific population, so that no certainty could be obtained on the possible changes in the study populations used in each of the publications. Not taking into account abstinence therefore introduced the possibility of confounding regardless of the recorded time-interval.

6.3 Complicating factors

It was decided not to include any judgment on the other confounders such as season of sampling, lifestyle, profession and diseases in the separate analyses because these factors were not addressed thoroughly in any of the publications, with most publications not addressing them at all. The fact that these were ignored in most of the articles could have exerted additional disturbing influence on the results.
Some of these confounders, such as season of sampling, could have easily been included in the studies. Still, season of sampling was only accounted for in the article by Fisch *et al.* (1997). For the other articles it is difficult to judge whether not recording or mentioning season of sampling introduced confounding. It has to be remembered however that season of sampling could be a factor of importance in the assessment of sperm parameter values (Levine 1991, Schrader *et al.* 1991, Fisch *et al.* 1997) and that when season of sampling is not recorded this could be masking a difference between the seasons of measurement that would introduce confounding.
The influence of stress on sperm motility can be substantial. A recent article by Fukuda *et al.* (1996) reported on the effect of experiencing an earthquake on sperm concentration and motility. They observed that acute stress resulting from a catastrophic earthquake caused reduced sperm motility, and even that time until recovery could be dependent on the severity of the damage, thus stress, experienced. This clearly indicates the important influence of stress on sperm quality. Correcting sperm measurements for this influence is hardly possible since stress in a subject (man) can have several causes and cannot be quantified.
Like stress, there are still other factors that influence sperm parameters and which are very difficult to correct for in analysis, like infections or exposure to radiation or specific agents. Therefore some uncertainties will always remain in sperm parameter measurements.

6.4 Trends in factors affecting sperm quality

Although no indications were found for an actual decline in sperm quality/quantity, the possibility exists that a decline did occur during the last decades as a result of changes in lifestyle and other confounders that occured in the normal population have occurred that could have caused a reduction in sperm quality /quantity over the last decades. Stress has possibly increased as a result of changed lifestyle and a steadily increasing population.
As for varicoceles, there is the possibility that their prevalence increased during that last decades. Increasing body length could increase the risk of varicocele and since people are growing larger each generation, this may influence the overall sperm quality of the total population. The fact that tall men run a higher risk of developping a varicocele may be related to a higher blood pressure in the testes. Indications for the relationship between length and varicoceles were observed in studies designed to investigate the influence of testosteron treatment on constitutional tall boys. Lemcke *et al.* (1996) noticed a higher prevalence of varicoceles and history of maldescent testis in the testosterone treated tall men compared with the controls. However, they did not examine whether this difference was due to the treatment or to tall statue alone. De Waal *et al.* (1995) too observed a high incidence of varicoceles (about 40%) with tall men. Moreover, they found the increased incidence of varicoceles to be unrelated to the treatment with testosterone.

During the last 50 years large numbers of men spent their time sitting in an office or driving a car (Forti & Serio 1993). It is probable that their number increased over the years, thereby increasing the number of men suffering from a reduced sperm quality through the deteriorating influence of increased temperature of the testes.
Environmental pollution increased over the last decades. It is very well possible that this had some effect on sperm quality or quantity. Sram *et al.* (1996) for example found clues for a damaging effect of air pollution in the Czech Republic on sperm motility and morphology.
These are all indications that a decline in sperm quality or quantity, if it actually occured, could be attributed to any of these or other still unknown factors and not only to environmental estrogens as proposed by Carlsen *et al.* (1992).

6.5 Fluctuations over the years

There are several indications for a fluctuation in sperm concentration over the years. This was observed by McLeod & Wang (1979) analyzing semen from 9000 men and by Fisch *et al.* (1997). The observed best-matching curves in the Greek study (Adamapoulos *et al.* 1996) seemed to indicate fluctuations as well. Although they found sperm count to have decreased over the study period, the observed values were in accordance with the fluctuations in Fisch *et al.* (1997), by starting the study in 1977 (high values) and ending it in 1993 (low values).
Therefore a chance exists that the trend observed in any study would not have been observed if more or other years would have been studied. This should always be kept in mind when conducting such an analysis. Possible explanations for the natural fluctuations are difficult to find (McLeod & Wang 1979). Fisch *et al.* (1997) proposed the influence of heat, another possible explanation might be the influence of environmental pollution.

6.6 Influence of geography and populations

The comparability of publications is reduced because measurements are obtained in different countries. There are good indications for geographical distinct sperm concentrations (see paragraph 4.2.3.1, Fisch *et al.* 1996). With regard to the other parameters, no information could be obtained. Whether these vary as well and whether different trends exist in the distinct countries is uncertain but very well possible.
Furthermore, measurements were obtained in different populations: men from infertile couples, potential sperm donors either fertile or with unknown fertility, and men banking sperm before vasectomy[1]. None of these, maybe with an exception for the potential sperm donors of unknown fertility, can be regarded as representative for the normal population (Sherins 1995). Changes in sperm parameters over time could be different in these populations, which can be measured by analyzing them separately. Yet, the meaning of such information is limited since these populations are often not well-defined or highly selected and because one is more interested in changes in sperm parameters concerning the general population. None of the articles that reported on men from infertile couples did mention their definition of an infertile couple. Although 'infertility' is not easy to define (Sheriff 1995, van Roijen *et al.* 1995), the WHO arbitrarily defined it as couples having 1 year of unprotected coitus without conceiving

[1] surgical removal of the vas deferens, or a portion of it.

(WHO 1992). Whether this definition is used in the articles is unknown. An additional question is whether a population of men with infertile marriage should be used at all because of the direct relationship between sperm parameters and the criterion for inclusion in this group.

6.7 Sperm quality and quantity versus fertility

The definition of infertility formulated by the WHO does not shed any light on the underlying relation with sperm quality or quantity. Although sperm concentrations below $20*10^6$/ml are regarded as abnormal (WHO 1992), even men with sperm concentrations below $5*10^6$/ml are observed to get women pregnant (Weber *et al.* 1995). This observation indicates that concentration *per se* is a poor index of fertility potential and suggests that functional sperm capacity is more important than sperm number. Male fertility potential therefore is very likely to be a multifactorial phenomenon depending upon a combination of different sperm parameters such as sperm count and, more important, the percentage of forms with progressive motility and normal morphology (Forti & Serio 1993, Oehninger & Kruger 1995, Morgentaler *et al.* 1995, Bollendorf *et al.* 1996).
Concentration is the only parameter that was measured in all publications, no doubt as a result of the relatively easy and objective assessment of this parameter.
Notwithstanding the fact that a decease in sperm count in the general population can influence fertility, studies should rather focus on changes in all parameters to be able to pronounce on the fertility potential of the general population. The problem is that the assessment of the qualitative parameters is much less objective than the quantitative ones. Internal quality control is therefore, as mentioned in paragraph 3.1.4, essential.

6.8 Ideal design of studies on possible changes in sperm quality/quantity

For a valid time-analysis of sperm parameters it is very important to correct for the confounders age, duration of abstinence and season of sampling. Furthermore, validation of the sperm analysis methodology should be performed regularly. The study population should be representative of the total population, which means no infertile men or sperm donors should be used as a population and no men should be excluded. It is impossible for retrospective studies to meet these criteria, as was observed in the analysis of the separate publications. The data used in these publications were collected for other purposes and therefore did not include essential information. Consequently, prospective studies are therefore necessary to assess the probability of an actual decline in sperm parameters. The major drawback of such prospective studies is that results will not become available until ten to twenty years from now and as such they are not useful for the present assessment of sperm quality or sperm quantity changes.

7. CONCLUSION

The meta-analysis on a change in sperm concentration during the last 50 years by Carlsen *et al.* (1992) suffers from various sources of bias and confounding. Therefore the evidence for a true decline in sperm concentration is not convincing.
Analysis of the separate epidemiological studies showed that, although the results were presented as good evidence for either a decline or no change, also here bias and confounding invalidate the conclusions. Actually, not one publication was found to be completely free of flaws but their severity varied among the articles. Both the articles that describe a downward trend in sperm parameters and the articles that observed no decline at all had similar flaws, so no general distinction could be made.
The following articles had a relatively good design: Auger *et al.* (1995), Adamapoulos *et al.* (1996), Paulsen *et al.* (1996) and Fisch *et al.* (1996). In these publications some uncertainties but no manifest errors occurred, suggesting that their results may be more reliable. Two of these publications observed a decline, one observed no change and one observed an increase in sperm concentration. All observed an unchanged seminal volume or even a slight increase. This may be taken as an indication that seminal volume has not changed during the last decades.
Therefore, the analysis of all articles analyzed revealed that the evidence to justify the conclusion that sperm quality or quantity has declined over the past decades is not convincing.

REFERENCES

Adami H, Bergström R, Möhner M, Zatonski W, Storm H, Ekbom A, Tretli S, Teppo L, Ziegler H, Rahu M, Gurevicius R & Stengrevics A. Testicular cancer in nine Northern European countries. *Int J Cancer* 1994;59:33-8

Adamopoulos DA *et al.* Seminal volume and total sperm number trends in men attending subfertility clinics in the greater Athens area during the period 1977-1993. *Hum Reprod* 1996;11(9):1936-41

Ansell PE *et al.* Cryptorchidism: a prospective study of 7500 consecutive male births, 1984-8. *Arch Dis Child* 1992;67:892-9

Auger J, Kunstmann J-M, Czyglik F, Jouannet P. Decline in semen quality among fertile men in Paris during the past 20 years. *N Engl J Med* 1995;332:281-5

Bahadur G, Ling KLE & Katz M. Statistical modelling reveals demography and time are the main contributing factors in global sperm count changes between 1938 and 1996. *Hum Reprod* 1996;vol 11 no 9:2635-39

Bendvold E. Semen quality in Norwegian men over a 20-year period. *Int J Fertil* 1989;34(6):401-4

Bendvold E, Gottlieb C *et al.* Depressed semen quality in Swedish men from barren couples: a study over three decades. *Arch Androl* 1991;26(3):189-94

Bergeron JM, Crews D, McLachlan JA. PCBs as environmental oestrogens: Turtle sex determination as a biomarker of environmental contamination. *Environ Health Prospect* 1994;102:780-1

Berman NG, Wang C, Paulsen. Methodological issues in the analysis of sperm concentration data. *J Androl* 1996;17:68-73

Bollendorf *et al.* Evaluation of the effect of the absence of sperm with rapid and linear progressive motility on subsequent pregnancy rates following intrauterine insemination or *in vitro* fertilization. *J Androl* 1996;17(5):550-7

Bonde JP *et al.* Identifying environmental risk to male reproductive function by occupational sperm studies: logistics and design options. *Occup and Envirnm Medicine* 1996;53:511-19

Bostofte E, Serup, Rebbe. Has The fertility of Danish men declined through the years in terms of semen quality? A comparison of semen qualities between 1952 and 1972. *Int J Fertil* 1983;28:91-5

Brake, Krause. Decreasing quality of semen. *Br Med J* 1992;305:1498

British Broadcasting Corporation. Horizon: Assault on the male. London: BBC, 1993

Bromwich P *et al.* Decline in sperm counts: an artefact of changed reference range of "normal"? *Br Med J* 1994;309:19-22

Bujan L, Mansat A, Pontonnier F, Mieusset R. Time series analysis of sperm concentration in fertile men in Toulouse, France. *Br Med J* 1996;312:471-2

Carlsen E, Giwercman A, Skakkebaek *et al.* Evidence for decreasing quality of semen during past 50 years. *Br Med J* 1992;305:609-13

Chia SE, Ong, Tsakok. Effects of cigarette smoke on human semen quality. *Arch Androl* 1994;33(3):163-8

Colborn Th, Dumanoski D & Peterson Meyers J. Our stolen future: are we threatening our fertility, intelligence, and survival? A scientific detective story. Dutton Books, New York, 1996

Comhaire F, Van Waeleghem K *et al.* Declining sperm quality in European men. *Andrologia* 1996;28:300-1
De Mouzon J, Spira A, Thonneau P, Multigner L (1996). Semen quality has declined among men born in France since 1950 [letter]. *Br Med J*;313:43.

Farrow S. Falling sperm quality: fact or fiction? *Br Med J* 1994;309:1-2

Figà-Talamanca I *et al.* Male infertility and occupational exposures: a case control study. *J Occupat Med Toxicol*;1:255-64

Fisch H *et al.* Semen analyses in 1,283 men form the United States over a 25-year period: no decline in quality. *Fertil Steril* 1996;65(5):1009-14

Fisch H, Andrews H *et al.* The relationship of sperm count to birth rates: a population based study. *J Urol* 1997;157:840-3

Fisch H, Goluboff ET. Geographic variations in sperm counts: a potential cause of bias in studies of semen quality. *Fertil Steril* 1996;65(5):1044-6

Forti G, Serio M. Male infertility: is its rising incidence due to better methods or detection or an increasing frequency? *Hum Reprod* 1993;8:1153-4

Fukuda M, Fukuda K *et al.* Kobe earthquake and reduced sperm motility. *Hum Reprod* 1996;11(6):1244-6

Gill WB, Schumacher GFB, Bibbo M, Straus FHI & Schoenberg HW. Association of diethylstilbestrol exposure in utero with cryptorchidism, testicular hypoplasia and semen abnormalities. *J Urol* 1979;122:36-9

Ginsberg J, Okolo *et al.* Residence in the London area and sperm density [letter]. *Lancet* 1994;343:230

Haidl G, Jung A, Schill WB. Ageing and sperm function. Hum Reprod 1996;11(3):558-60

Henderson BE, Benton B, Cosgrove M, Baptista J, Aldrich J, Townsend D, Hart W & Mack TM. Urogenital tract abnormalities in sons of women treated with diethylstilbestrol. *Pediatrics* 1976;58:505-7

Hennekens CH & Buring JE. Epidemiology in medecine. Boston/Toronto: Little, Brown and company, 1987

de Hollander AEM, Preller EA, Heisterkamp SH & Jansen J. Meta-analyse van observationeel onderzoek. Mogelijkheden en beperkingen bij toepassing ten behoeve van het kwanitficeren van gezondheidsrisico's. Bilthoven: RIVM, rapportnr. 263610 002, 1996

Holzki G, Gall H, Hermann J. Cigarette smoking and sperm quality. *Andrologia* 1991;23(2):141-4

Irvine DS. Falling sperm quality [letter]. *Br Med J* 1994;309:476

Irvine DS *et al.* Evidence of deteriorating semen quality in the United Kingdom: birth cohort study in 577 men in Scotland over 11 years. *Br Med J* 1996;312:467-71

James WH. Secular trend in reported sperm counts. *Andrologia* 1980;12:381-8

James WH. The decline in sex ratios at birth, England and Wales, 1973-1990. *J Epidem Community Health* 1996;50(6):690-1

Jequier AM, Ukombe EB. Br J Urol 1983;55:434-6

Joffe M. Recent decline may be relatively late stage of long term proces [letter]. BMJ 1996;131:44

Keiding N, Giwecmann A *et al.* Falling sperm quality. *Br Med J* 1994;309:131

Kruger TF, Menkveld R, Stander FSH, Lombard CJ, Van der Merwe JP, van Zyl JA et al. Sperm morphologic features as a prognostic factor in in vitro fertilization. Fertil Steril 1986;46:1118-23

Le Lannou D & Griveau J-F. Dialy sperm production and daily output in men. In: Hamamah S & Mieusset R, Research in the male gametes: production and quality. INSERM 1996

Lemcke B, Zentgraf J *et al.* Long-term effects on testicular function of high-dose testosterone treatement for excessively tall stature. *J Clin Endocrinol Metab* 1996;81(1):296-301

Leto S, Frensilli FJ. Changing parameters of donor semen. *Fertil Steril* 1981;36:766-70

Levine RJ. Seasonal variation in human semen quality. *Adv Exp Med Biol* 1991;286:89-96

MacLeod J, Gold RZ. The male factor in fertility and infertility. II. Spermatozoon counts in 1000 men of known fertility and in 1000 cases of infertile marriage. *J Urology* 1951;66:436-49

MacLeod J, Wang Y. Male fertility potential in terms of semen quality: a review of the past, a study of the present. *Fertil Steril* 1979;31:103-16

Magnus O, Tollefsrud A *et al.* Effects of varying the abstinence period in the same individuals on sperm quality. *Arch Androl* 1991;26(3):199-203

Mallidis C, Howard EJ, Baker HW. Variation of semen quality in normal men. *Int J Androl* 1991;14(2):99-107

Matlai P & Beral V. Trends in congenital malformations of external genitalia. *Lancet* 1985;i:108

Matson PL. External assessment for semen analysis and sperm antibody detection: results of a pilot scheme. *Hum Reprod* 1995;vol 10 no 3:620-25

Menchini-Fabris F, Rossi P, Palego P, Simi S, Turchi P. Declining sperm counts in Italy during the past 20 years. *Andrologia* 1996;28;304

Menkveld R *et al.* Possible changes in male fertility over a 15-year period. *Arch Androl* 1986;17:143-4

Mieusset R, Bujan L, et al. Association of scrotal hyperthermia with impaired spermatogenesis in infertile men. Fertil Steril 1987;48(6):1006-11

Morgenthaler A, Fung, Harris *et al.* Sperm morphology and in vitro fertilisation outcome: a direct comparison of WHO and strict criteria methodologies. *Fertil Steril* 1995;64(6):1177-82

Mortimer D. The male factor in infertility. Part I. Semen analysis. *Curr Probl Obstet Gynecol Fertil* 1985;8(7):3-87

Mortimer D, Shu MA, Tan R. Standardization and quality control of sperm concentration and sperm motility counts in semen analysis. Hum Reprod 1986;1(5):299-303

Naghma-E-Rehan *et al.* The semen of fertile men. Statistical analysis of 1300 men. *Fertil Steril* 1975;26:492-502

Nederlands Dagblad. Stof in kindervoeding remt voortplanting. 28/05/1996

Neuwinger J, Behre, Nieschlag. External quality control in the andrology laboratory: an experimental multicenter trial. *Fertil Steril* 1990;54:308-14

Nnatu SN, Giwa-Osagie OF, Essien EE. Effect of repeated semen ejaculation on sperm quality. *Clin Exp Obstet Gynecol* 1991;18(1):39-42

Oehninger S, Kruger T. The diagnosis of male infertility by semen quality. Clinical significance of sperm morphology assessment. *Hum Reprod* 1995;10(5):1037-8

Oldereid NB *et al.* Life styles of men in barren couples and their relationship to sperm quality. *Int J Fertil* 1992;37(6):343-9

Olsen J. Is human fecundity declining-and does occupational exposures play a role in such a decline if it exists? *Scand J Work Environ H* 1994;20(special issue):72-7

Osser S, Liedholm P, Ranstam J. Depressed semen quality: a study over two decades. *Arch Androl* 1984;12:113-6

Paduch DA, Niedzielski J. Semen analysis in young men with varicocele: preliminary study. *J Urology* 1996;156:788-90

Parool. Bestrijdingsmiddelen riskanter dan gedacht. 07/06/1996

Paulsen CA, Berman NG, Wang C. Data from men in greater Seattle area reveals no downward trend in semen quality: further evidence that deterioration of semen quality is not geographically uniform. *Fertil Steril* 1996;65(5):1015-20

Pellestor E, Girardet A, Andreo B. Effect of long abstinence periods on human sperm. *Int J Fertil* 1994;39(5):278-82

Pol PS *et al.* Circannual rhythm of sperm parameters of fertile men. *Fertil Steril* 1989;51:1030-3

Purdom CE *et al.* Estrogenic effects of effluents from sewage treatement works. *Chem Ecol* 1994;8:275-85

van Roijen JH *et al.* Standaardisatie van semen-analyse. *Ned Tijdschr Klin Chem* 1995;20:209-12

Schrader SM, Turner TW, Simon SD. Longitudinal study of semen quality of unexposed workers. Sperm motility characteristics. *J Androl* 1991;12(2):126-31

Schwartz D *et al.* Semen characteristics as a function of age in 833 fertile men. *Fertil Steril* 1983;39:530-5

Seracchioli R *et al.* The diagnosis of male infertility by semen quality. Sperm morphology is not the only criterion of male infertility. *Hum Reprod* 1995;10(5):1039-41

Sharpe RM. Declining sperm counts in men: is there an endocrine cause? *J Endocrinol* 1993;136:357-60

Sheriff DS. Setting standards of male fertility. I. Semen analysis in 1500 patients-a report. *Andrologia* 1983;15:687-92

Sheriff DS. Analysis of semen in a constantly changing social context of medicine. *Arch Androl* 1995;34(5):125-32

Sherins RJ, Brightwell D & Sternthal M. Longitudinal analysis of semen of fertile and infertile men. In: The testis in normal and infertile men, ed. Troen P & Nankin HR. Raven Press, New York 1977

Sherins RJ. Are semen quality and male fertility changing? *N Engl J Med* 1995;332(5):327-8

Smith KD & Steinberger E. What is oligospermia? In: The testis in normal and infertile men. Ed.Troen P & Nankin HR. Raven Press, New York 1977

Sram IB *et al.* Teplice programm--The impact of air pollution on human health. *Environ Health Persp* 1996;104, Suppl 4:699-714

Stillman RJ. In utero exposure to diethylstilbestrol: adverse effects on the reproductive tract and reproductive performance in male and female offspring. *Am J Obstet Gynecol* 1982;142:905-21

Tas S *et al.* Occupational hazards for the male Rproductive system. *Critical Reviews in Toxicology* 1996;26(2):261-307

Tielemans E, Heederik D *et al.* Intraindividual variability and redundancy of semen parameters. *Epidemiology* 1997;8(1):99-103

Tjoa WS *et al.* Circannual rhythm in human sperm count revealed by serially independent sampling. *Fertil Steril* 1982;38:454-9

Toppari *et al.* Male reproductive health and environmental chemicals with estrogenic effects. Kopenhagen, Denemarken: Miljø- og Energiministeriet MILJØstyrelsen. Miljøproject 290, 1995.

Tyler JP, Crockett, Driscoll. Studies of human seminal parameters with frequent ejaculation. I. Clinical characteristics. *Clin Reprod Fertil* 1982;1:273-85

Ulstein M. Semen quality--has it changed during the last decades? *Acta Obstet Gynecol Scand* 1996;75(3):201-2

Vine MF. Worldwide decline in semen quality might be due to smoking. *Br Med J* 1996;312:506

Virula M, Niemi M *et al.* High and unchanged sperm counts of Finnish men. *Int J Androl* 1996:19(1):11-7

de Waal WJ, Vreeburg JTM *et al.* High dose testosteron therapy for reduction of final height in constitutional tall boys: does it influence testicular function in adulthood? *Clinical Endocr* 1995;43:87-95

Waeleghem K van *et al.* Deterioration of sperm quality in young healthy Belgian men. *Human Reprod* 1996;11:325-9

Wang C *et al.* Cross-sectional study of semen parameters in a large group of normal Chinese men. *Int J Androl* 1985;8:257-74

Weber RFA *et al.* De rol van andrologie bij diagnostiek en behandeling van fertiliteitsstoornissen. *Ned Tijdschr Geneeskd* 1995;139(18):922-5

Wittmaack FM, Shapiro SS. Longitudinal study of semen quality in Winsconsin men over a decade. *Winsconsin Med J* 1992;91:91-5

World Health Organization (1980). *WHO Laboratory Manual for the Examination Of Human Semen and Semen-Cervical Mucus Interaction.* Press Concern, Singapore.

World Health Organization (1992). *WHO Laboratory Manual for the Examination Of Human Semen and Semen-Cervical Mucus Interaction.* Cambridge University Press, Cambridge.