Results of Proficiency Test
Formaldehyde in leather
October 2013

Organised by: Institute for Interlaboratory Studies
Spijkenisse, the Netherlands

Author: ing. R.J. Starink
Correctors: dr. R. Visser & ing. N. Boelhouwer
Report: iis13A04L

December 2013

CONTENTS

Appendices:

1 INTRODUCTION

Worldwide, many consumer products are produced from leather. During the production of leather products many different types of auxiliary agents and dyes are used to process leather. Neither in the U.S. nor in the European Union there is general legislation that limits the presence of formaldehyde in leather. However, some individual countries have restricting limits on the concentration of free formaldehyde in leather that may vary from 20 mg/kg for leather used for young children to 100 mg/kg when the leather is in contact with the skin, 150 mg/kg for shoe uppers and 400 mg/kg for leather without permanent contact with the skin. In 2006, The China Leather Industry Standard Committee Organization established the Limit of Harmful Matters in Leather: GB20400-2006. This national mandatory standard was approved by the General Administration of P.R. of China for Quality Supervision and Inspection and Quarantine and implemented in December 2007.
Since several years, the Institute for Interlaboratory Studies (iis) organises a proficiency scheme for Formaldehyde in textile. On request of several participants, the institute decided to organize also a proficiency test for Formaldehyde in Leather as part of the proficiency testing program 2013/2014.
In this first interlaboratory study 49 laboratories in 17 different countries participated. See appendix 2 for the number of participating laboratories per country.
In this report, the results of this 2013 proficiency test are presented and discussed.

2 SET UP

The Institute for Interlaboratory Studies in Spijkenisse was the organiser of this proficiency test. Analyses of fit for use and homogeneity were subcontracted. In this proficiency test, it was decided to use one sample (#13192, approx. 3 grams) which was positive on Free Formaldehyde. Participants were requested to report results with one extra figure. These unrounded results were preferably used for the statistical evaluations.

2.1 QUALITY SYSTEM

The Institute for Interlaboratory Studies in Spijkenisse, the Netherlands, has implemented a quality system based on IEC/ISO17043:2010. This ensures strict adherence to protocols for sample preparation and statistical evaluation and 100% confidentiality of participant's data. Feedback from the participants on the reported data is encouraged and customer's satisfaction is measured on regular basis by sending out questionnaires.

2.2 PROTOCOL

The protocol followed in the organisation of this proficiency test was the one as described for proficiency testing in the report 'iis Interlaboratory Studies: Protocol for the Organisation, Statistics and Evaluation' of January 2010 (iis-protocol, version 3.2) which can be downloaded from http://www.iisnl.com.

2.3 CONFIDENTIALITY STATEMENT

All data presented in this report must be regarded as confidential and for use by the participating companies only. Disclosure of the information in this report is only allowed by means of the entire report. Use of the contents of this report for third parties is only allowed by written permission of the Institute for Interlaboratory Studies. Disclosure of the identity of one or more of the participating companies will be done only after receipt of a written agreement of the companies involved.

2.4 SAMPLES

A leather sample was shreddered and after homogenisation divided over 78 subsamples of approx. 3 gram and labelled sample #13192.
The homogeneity of the subsamples was checked on Formaldehyde according to ISO17226:08 on 8 randomly selected samples. The homogeneity testing was performed by a subcontracted ISO17025 accredited laboratory. See the following tables for the test results.

	Formaldehyde in mg/kg
Sample #13192-1	141.6
Sample #13192-2	140.1
Sample #13192-3	132.4
Sample #13192-4	146.8
Sample #13192-5	137.7
Sample #13192-6	131.8
Sample #13192-7	137.6
Sample #13192-8	138.1

Table 1: homogeneity test results of subsamples #13192

From the above test results, the repeatabilities were calculated and compared with 0.3 times the corresponding reproducibilities in agreement with the procedure of ISO 13528 (Annex B2) or with the repeatability of the reference method, in the next table:

	Formaldehyde in mg/kg Sample #13192
r	13.6
Reference test	ISO17226:08
$0.3*R_{(reference\ test)}$	19.3

Table 2: repeatability of subsamples #13192

From the above results of the homogeneity test, the repeatability was calculated.
The calculated repeatability for sample #13192 is in good agreement with 0.3 times the reproducibility of the reference test method. Therefore, homogeneity of all subsamples was assumed.

One sample of approx. 3 grams (labelled #13192) was sent to the participating laboratories on October 9, 2013.

2.5 ANALYSES

The participants were asked to determine on sample #13192, the concentration of Free Formaldehyde and the concentration of Formaldehyde emissions. To get comparable results, detailed report forms were sent together with each set of samples. On the report from the requested Formaldehyde concentrations, including the units was pre-printed. Also a letter of instructions was sent along.

3 RESULTS

During four weeks after sample despatch, the results of the individual laboratories were gathered. The original data are tabulated in the appendices of this report. The laboratories are presented by their code numbers.

Directly after the deadline, a reminder fax was sent to those laboratories that had not yet reported. Shortly after the deadline, the available results were screened for suspect data. A result was called suspect in case the Huber Elimination Rule (a robust outlier test, see lit.5) found it to be an outlier. The laboratories that produced these suspect data were asked to check the results. Additional or corrected data are placed under 'Remarks' in the result tables in appendix 1. A list of abbreviations used in the tables can be found in appendix 3.

3.1 STATISTICS

Statistical calculations were performed as described in the report 'iis Interlaboratory Studies: Protocol for the Organisation, Statistics and Evaluation' of January 2010 (iis-protocol, version 3.2).
For the statistical evaluation the *unrounded* (when available) figures were used instead of the rounded results. Results reported as '<…' or '>…" were not used in the statistical evaluation.
Before further calculations, the normality of the distribution of the various data sets per determination was checked by means of the Lilliefors-test. In the case of an anormal distribution, the statistical evaluation should be used with care.

According to ISO 5725 (1986 and 1994, lit.7 and 8) the original results per determination were submitted subsequently to Dixon's and Grubbs' outlier tests. Outliers are marked by D(0.01) for the Dixon's test, by G(0.01) or DG(0.01) for the Grubbs' test. Stragglers are marked by D(0.05) for the Dixon's test, by G(0.05) or DG(0.05) for the Grubbs' test. Both outliers and stragglers were not included in the calculations of averages and standard deviations. Finally, the reproducibilities were calculated from the standard deviations by multiplying them with a factor of 2.8.

For each assigned value the uncertainty was determined in accordance with ISO13528. Subsequently the calculated uncertainty was evaluated against the respective requirement based on the target reproducibility in accordance with ISO13528. When the uncertainty passed the evaluation no remarks are made in the report. However, when the uncertainty failed the evaluation it is mentioned in the report and it will have consequences for the evaluation of the test results.

3.2 GRAPHICS

In order to visualise the data against the reproducibilities from literature, Gauss plots were made, using the sorted data for one determination (see appendix 1). On the Y-axis the reported analysis results are plotted. The corresponding laboratory numbers are under the X-axis. The straight horizontal line presents the consensus value (a trimmed mean). The four striped lines, parallel to the consensus value line, are the +3s, +2s, -2s and -3s target reproducibility limits of the selected standard. Outliers and other data, which were excluded from the calculations, are represented as a cross. Accepted data are represented as a triangle. Furthermore, Kernel Density Graphs were made. This method is producing a smooth density approximation to a set of data that avoids some problems associated with histograms (see appendix 3; nos.14 and 15).

3.3 Z-SCORES

To evaluate the performance of the individual participating laboratories the z-scores were calculated.
In order to be able to have an objective evaluation of the performance of the individual participants, it was decided to evaluate this performance against the literature requirements. Therefore, the z-scores were calculated using a target standard deviation. This target standard deviation was calculated from the literature reproducibility by division with 2.8.

The $z_{(target)}$-scores were calculated according to:

$z_{(target)}$ = (individual result - average of proficiency test) / target standard deviation

The $z_{(target)}$-scores are listed in the result tables in appendix 1.

Absolute values for z<2 are very common and absolute values for z>3 are very rare. The usual interpretation of z-scores is as follows:

When a laboratory did use a test method with a reproducibility that is significantly different from the reproducibility of the reference test method used in this report, it is strongly advised to recalculate the z-score, while using the reproducibility of the actual test method used this in order to evaluate the fit-for-useness of the reported test result.

Absolute values for z<2 are very common and absolute values for z>3 are very rare. The usual interpretation of z-scores is as follows:

	| z | < 1	good
1 <	| z | < 2	satisfactory
2 <	| z | < 3	questionable
3 <	| z |	unsatisfactory

4 EVALUATION

During the execution of this proficiency test no problems occurred with the delivery of the samples. Two laboratories did not report any test results and five other laboratories reported results after the final reporting date.

Finally, the 48 reporting laboratories send in total 52 numerical results. Observed were 6 statistical outlying results, which is 11.5% of the numerical results. In proficiency studies, outlier percentages of 3% - 7.5% are quite normal.

For sample #13192, not normal distributions were found. Therefore the statistical evaluation for this sample should be used with due care.

4.1 EVALUATION PER SAMPLE

In this section, the sample #13192 is discussed. All statistical results reported on the leather sample are summarised in appendix 1.

Formaldehyde content (HPLC): This determination was not problematic. Only one statistical outlier was observed, but four test results were excluded from the statistical evaluation as the reported test method was intended for textiles only. The calculated reproducibility after rejection of the statistical outlier is in full agreement with the requirements of ISO17226-1:2008.

Formaldehyde content (colorimetric): This determination was very problematic. Five statistical outliers were observed. The calculated reproducibility after rejection of the statistical outliers is not at all in agreement with the requirements of ISO17226-2:2008.

Formaldehyde by Emission: None of the participants reported a test result for this determination. Therefore no conclusions were drawn.

4.2 PERFORMANCE EVALUATION FOR THE GROUP OF LABORATORIES

A comparison has been made between the calculated reproducibilities estimated from ISO17226 and the reproducibilities as found for the group of participating laboratories. The number of significant results, the average results, the calculated reproducibilities (standard deviation*2.8) and the target reproducibilities (ISO17226), are compared in the next table.

Parameter	unit	n	average	2.8 * sd	R (target)
Formaldehyde (HPLC)	mg/kg	23	72.4	44.6	45.6
Formaldehyde (colorimetric)	mg/kg	19	164.6	115.4	46.1

Table 3: reproducibilities of textile samples #13192

From the above tables it can be concluded that, without statistical calculations, the group of participating laboratories does not have any difficulties with the HPLC analysis, but severe problems with the colorimetric analysis, when compared with the requirements of the target test methods. See also the discussions in paragraphs 4.1 and 6.

5 DISCUSSION

The standard test method ISO17226 for determination of the formaldehyde is available. Part 3 of ISO17226 describes the determination of Formaldehyde emission. None of the participants used this test method.
Part 1 and part 2 describe the determination of the formaldehyde content by extraction of the free formaldehyde from the leather with a detergent solution. The difference between both parts of ISO17226 is the method of quantification. Quantification of the formaldehyde is done by HPLC in part 1 and by colorimetric analysis in part 2. Therefore part 2 is not selective for formaldehyde, whereas part 1 is selective. The test results from part 2 will in general be higher than the test results from part 1. In the case of dispute part 1 shall be used in preference.

When the results of this interlaboratory study were compared to the Standard “Limit of Harmful Matters in Leather” of the Chinese Leather Industry Committee Organization: GB20400-2006 (table 4), it may be noticed that not all participants would make identical decisions about the acceptability of the leather.

GB20400	Category A Products for babies: underclothes, bedding, etc	Category B Products with **Direct** skin contact	Category C Products Without direct skin contact
Free Formaldehyde in mg/kg	<20	<75	<300

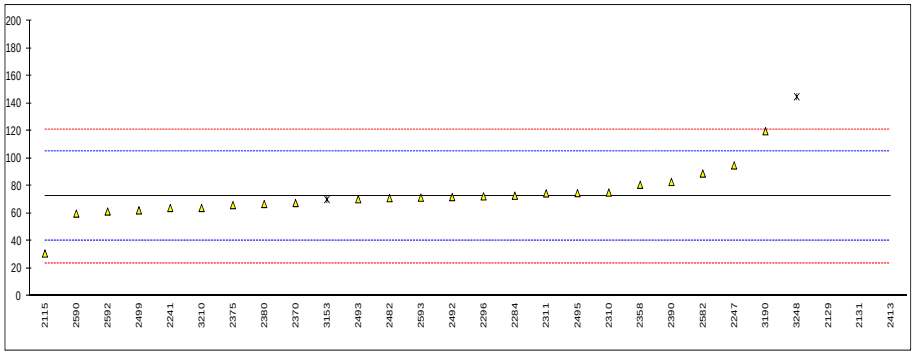
Table 4: Standard GB20400:2006

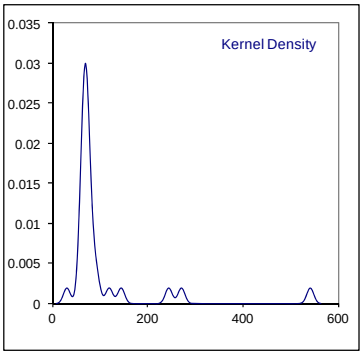
When using ISO17226 part 1, all reporting laboratories (28) would reject this sample for category A, ten laboratories would reject this sample also for category B and only one laboratory would also reject this sample for category C.
When using ISO17226 part 2, all reporting laboratories (24) would reject this sample for category A and category B. Five laboratories would also reject this sample for category C.

APPENDIX 1

Determination of Formaldehyde content (HPLC) on sample #13192; results in mg/kg

lab	method	value	mark	z(targ)	remarks
110		-----		-----	
2115	ISO17226-1	30.85		-2.55	
2129	ISO14184-1	244.9	ex	10.59	Result excluded, test method is for textile
2131	JIS LAW 112	271.65	ex	12.23	Result excluded, test method is for textile
2172		-----		-----	
2213		-----		-----	
2235		-----		-----	
2241	ISO17226-1	63.97		-0.52	
2247	ISO17226-1	95.0		1.39	
2273		-----		-----	
2284	ISO17226-1	72.91		0.03	
2295		-----		-----	
2296	ISO17226-1	72.46		0.00	
2300		-----		-----	
2310	ISO17226-1	75.2		0.17	
2311	ISO17226-1	74.6		0.13	
2358	ISO17226-1	80.869		0.52	
2370	ISO17226-1	67.58		-0.30	
2375	ISO17226-1	66.07		-0.39	
2380	ISO17226-1	66.89		-0.34	
2390	ISO17226-1	82.94		0.65	
2413	ISO17226-1	542.41	G(0.01)	28.86	
2452		-----		-----	
2459		-----		-----	
2472		-----		-----	
2482	ISO17226-1	71.125		-0.08	
2492	ISO17226-1	71.9		-0.03	
2493	ISO17226-1	70.33		-0.13	
2495	ISO17226-1	74.8		0.15	
2499	GB/T19941	62.26		-0.62	
2518		-----		-----	
2519		-----		-----	
2573		-----		-----	
2575		-----		-----	
2578		-----		-----	
2582	ISO17226-1	89.059		1.02	
2585		-----		-----	
2586		-----		-----	
2590	ISO17226-1	59.89		-0.77	
2592	ISO17226-1	61.41		-0.68	
2593	ISO17226-1	71.47		-0.06	
3100		-----		-----	
3153	ISO14184-1	70.16	ex	-0.14	Result excluded, test method is for textile
3190	ISO17226-1	119.85		2.91	
3209		-----		-----	
3210	ISO17226-1	64.0		-0.52	
3220		-----		-----	
3237		-----		-----	
3248	GB/T2912-1	145	ex	4.46	Result excluded, test method is for textile
	normality	not OK			
	n	23			
	outliers	1	+ 4 excl		
	mean (n)	72.41			
	st.dev. (n)	15.940			
	R(calc.)	44.63			
	R(ISO17226-1)	45.60			

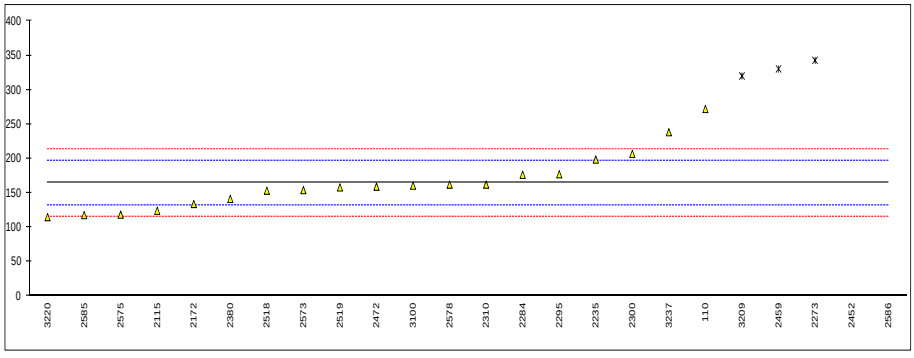


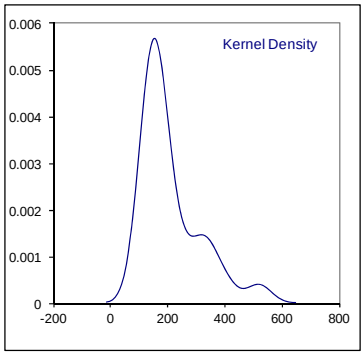


Determination of Formaldehyde content (Colorimetric) on sample #13192; results in mg/kg

lab	method	value	mark	z(targ)	remarks
110	ISO17226-2	272.1		6.53	
2115	ISO17226-2	123.88		-2.47	
2129		-----		-----	
2131		-----		-----	
2172	ISO17226-2	133.76		-1.87	
2213		-----		-----	
2235	ISO17226-2	198.49		2.06	
2241		-----		-----	
2247		-----		-----	
2273	ISO17226-2	343.00	G(0.05)	10.84	
2284	ISO17226-2	176.38		0.72	
2295	ISO17226-2	177		0.75	
2296		-----		-----	
2300	ISO17226-2	206.84		2.57	
2310	ISO17226-2	162		-0.16	
2311		-----		-----	
2358		-----		-----	
2370		-----		-----	
2375		-----		-----	
2380	ISO17226-2	141.18		-1.42	
2390		-----		-----	
2413		-----		-----	
2452	ISO17226-2	400.7	C,G(0.01)	14.34	First reported 302.1
2459	ISO17226-2	330.5	G(0.05)	10.08	
2472	ISO17226-2	159.02		-0.34	
2482		-----		-----	
2492		-----		-----	
2493		-----		-----	
2495		-----		-----	
2499		-----		-----	
2518	ISO17226-2	153.1705		-0.69	
2519	ISO17226-2	157.9		-0.41	
2573	ISO17226-2	153.92		-0.65	
2575	ISO17226-2	118.26	C	-2.81	First reported 393.59
2578	ISO17226-2	161.99		-0.16	
2582		-----		-----	
2585	ISO17226-2	117.54		-2.86	
2586	ISO17226-2	519.11	C,G(0.05)	21.53	First reported 817.08
2590		-----		-----	
2592		-----		-----	
2593		-----		-----	
3100	ISO17226-2	160.49		-0.25	
3153		-----		-----	
3190		-----		-----	
3209	ISO17226-2	320.11	G(0.05)	9.45	
3210		-----		-----	
3220	ISO17226-2	114.67		-3.03	
3237	ISO17226-2	238.31		4.48	
3248		-----		-----	

normality	not OK
n	19
outliers	5
mean (n)	164.57
st.dev. (n)	41.217
R(calc.)	115.41
R(ISO17226-2:08)	46.11





APPENDIX 2

Number of participants per country

1 lab in BANGLADESH
1 lab in BULGARIA
1 lab in FRANCE
2 labs in GERMANY
5 labs in HONG KONG
1 labs in HUNGARY
6 labs in INDIA
7 labs in ITALY
13 labs in P.R. of CHINA
2 labs in PAKISTAN
1 lab in SRI LANKA
1 lab in SWITZERLAND
1 lab in TAIWAN R.O.C.
1 lab in TUNISIA
3 labs in TURKEY
2 labs in U.S.A.
1 lab in UNITED KINGDOM

APPENDIX 3

Abbreviations:

C = final result after checking of first reported suspect result
D(0.01) = outlier in Dixon's outlier test
D(0.05) = straggler in Dixon's outlier test
G(0.01) = outlier in Grubbs' outlier test
G(0.05) = straggler in Grubbs' outlier test
DG(0.01) = outlier in Double Grubbs' outlier test
DG(0.05) = straggler in Double Grubbs' outlier test
n.a. = not applicable
n.d. = not detected
W = withdrawn

Literature:

1. iis Interlaboratory Studies, Protocol for the Organisation, Statistics & Evaluation, January 2010
2. GB20400-2006 Standard;.
3. Thai Green label. TGL-16. July 2002.
4. Impacts of Environmental Standards and requirements in EU Countries. Aug 99.
5. Horwitz. Journal of AOAC International Vol. 79 No.3. 1996.
6. P.L. Davies. Fr Z. Anal. Chem. 351. 513 (1988).
7. W.J. Conover. Practical; Nonparametric Statistics. J. Wiley&Sons. NY. p.302 (1971).
8. ISO 5725. (1986).
9. ISO 5725. parts 1-6 (1994).
10. ISO17226-1: 2008.
11. ISO17226-2: 2008.
12. ISO17226-3: 2008
13. M. Thompson and R. Wood. J. AOAC Int. 76. 926. (1993).
14. Analytical Methods Committee Technical brief, No4 January 2001.
15. The Royal Society of Chemistry 2002, Analyst 2002, 127 page 1359-1364, P.J. Lowthian and M. Thompson.(see http://www.rsc.org/suppdata/an/b2/b205600n/)
16. Official Journal of the European Communities L133/29: May 2002.