Available online at www.sciencedirect.com
ScienceDirect
Chemosphere 69 (2007) 1287–1294
CHEMOSPHERE
www.elsevier.com/locate/chemosphere

Agreement between breast milk dioxin levels by CALUX bioassay and chemical analysis in a population survey in Hong Kong

L.L. Hui [a], A.J. Hedley [a,*], E.A.S. Nelson [b], R. Malisch [c], T.W. Wong [d], B.J. Cowling [a]

[a] *Department of Community Medicine, The University of Hong Kong, 5/F, William M.W. Mong Block, 21 Sassoon Road, Pokfulam, Hong Kong*
[b] *Department of Paediatrics, The Chinese University of Hong Kong, 6/F Clinical Sciences Building, Prince of Wales Hospital, Shatin, Hong Kong*
[c] *The State Institute for Chemical and Veterinary Analysis of Food, Bissierstr. 5, D-79114 Freiburg, Germany*
[d] *Department of Community and Family Medicine, The Chinese University of Hong Kong, 4/F School of Public Health, Prince of Wales Hospital, Shatin, Hong Kong*

Received 8 February 2007; received in revised form 14 May 2007; accepted 15 May 2007
Available online 5 July 2007

Abstract

Chemically-activated luciferase gene expression (CALUX) bioassay and gas chromatography/mass spectrometry (GC/MS) are used to determine dioxin levels in food and humans. Valid measures of the agreement between the two methods would improve interpretation of bioassay results. Paired breast milk samples from 250 mothers, as 11 pooled samples, were analysed by GC/MS for total WHO-TEQ (7 polychlorinated dibenzo-*para*-dioxins, 10 polychlorinated dibenzofurans and 12 dioxin-like polychlorinated biphenyls) and as individual samples by CALUX. Mean difference between total WHO-TEQ (weighted by TEF system derived in 1997) and mean CALUX-TEQ in each pool was 1.6 pg/g fat (95% CI: 0.7, 2.4), indicating a statistically significant overestimation of CALUX-TEQ compared to WHO-TEQ, probably due to the presence of Ah-receptor agonists. CALUX estimated toxicity of 13 pg/g fat was greater than the WHO-TEQ by 0.9, 3.1 and 0.3 pg/g fat for mothers from Hong Kong, mainland China and overseas territories, respectively. When the 2005 TEF system was applied, a reduction of 14–26% in the WHO-TEQ and a larger but less disperse discrepancy between WHO-TEQ and CALUX-TEQ (3.9 pg/g fat, 95% CI: 3.5, 4.4) were observed. Our study suggested that the mothers' place of residence explained the discrepancy between CALUX-TEQ and WHO-TEQ and should be considered in inter-country comparisons for CALUX-TEQ. For regulatory purposes bioassays for detecting quantitative dioxin contents in any setting must be combined with adequate extraction, clean-up and validation with WHO-TEQs. The larger difference between the two measurements after using the new TEF system warrants further investigation.

Keywords: Dioxins; Chemical determination; Bioassay; Toxic equivalent factor; Breast milk; Hong Kong

1. Introduction

Levels of polychlorinated dibenzo-*p*-dioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs) have been determined in humans since the 1970s. The accepted standard method of estimation for global population surveys and regulatory purposes is based on high-resolution gas chromatography/high-resolution mass spectrometry (GC/MS). An established convention in this chemo-analytical method is to report their concentrations in units of "2,3,7,8-tetra-chlorodibenzo-*para*-dioxin (TCDD) toxic equivalents" or TEQ, the sum of the concentrations of the target congeners, 17 2,3,7,8-substituted PCDD/Fs and 12 dioxin-like polychlorinated biphenyls (PCBs), weighted according to their toxicity by toxic equivalence factors (TEF) relative to the most toxic congener TCDD. (van den Berg et al., 1998; van den Berg et al., 2006).

Abbreviations: PCDD, polychlorinated dibenzo-*para*-dioxins/dioxins; PCDF, polychlorinated dibenzofurans/furans; PCB, polychlorinated biphenyls; WHO, World Health Organization; GC/MS, gas chromatography with mass spectrometry; CALUX, chemically activated luciferase gene expression; TEQ, toxic equivalent; TEF, toxic equivalence factors; Ah, aryl hydrocarbon; SD, standard deviation; TCDD, 2,3,7,8-tetra-chlorodibenzo-*para*-dioxin.

* Corresponding author. Tel.: +852 28199282; fax: +852 28559528.
E-mail address: commed@hkucc.hku.hk (A.J. Hedley).

0045-6535/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chemosphere.2007.05.038

The application of GC/MS and the TEF system is considered to be the most valid method or "gold standard" for determining the toxicity of dioxin-like compounds in the environment. Biotechnology has been applied to the estimation of dioxin toxicity (Behnisch et al., 2001a), by measuring biological responses such as enzyme activity (Safe, 1993), expression of reporter genes (Garrison et al., 1996), ligand–receptor binding (Seidel et al., 2000) or antigen–antibody reaction (Harrison and Eduljee, 1999), to the unique structural properties of dioxin-like compounds. These bioanalytical detection methods such as CALUX (chemically activated luciferase gene expression), aryl hydrocarbon (Ah) immunoassays, EROD (7-ethoxyresorufin-*O*-deethylase) and enzyme immunoassays have been used for rapid screening of various matrices such as food, sediments, soil and fly ash. (JECFA, 2002) A reporter gene assay [Method 4425] and an immunoassay [Method 4025] have been approved by the United States Environmental Protection Agency for screening extracts of environmental samples for planar organic compounds and dioxins respectively. Bioassay provides information about toxicological effects (Behnisch et al., 2001b) and are potentially more efficient and economical in assessing environmental health risks than chemical methods. However, the European Commission Directives required that positive results of these bioanalytical detection methods for food stuff (European Commission, 2002a) and feeding stuff (European Commission, 2002b) should be confirmed by GC/MS analysis.

The determination of dioxin toxic equivalents is based on the assumption that all dioxin-like compounds act through the Ah-receptor signal transduction pathway. It has been shown that the strength with which congeners bind to the Ah-receptor is directly proportional to the toxicity, enhanced gene transcription and enzyme activities mediated by the Ah-receptor mechanism (Giesy et al., 2002). Recently Windal et al. (2005) reviewed the performance of CALUX as a bioanalytical tool for estimation of dioxin-like toxicity and examined parameters which might affect the quality and accuracy of the analysis. They pointed to a number of parameters, including clean-up procedures, use of solvents and extracts and their exposure duration, and types of cell-line, which may all affect the result of CALUX. The presence of Ah-receptor agonists in the mixtures might lead to false positive results in Ah-receptor based bioassays (van Wouwe et al., 2004a). False negative results are also possible if toxic effects of dioxins and related compounds are Ah-receptor independent (Butler et al., 2004) or possibly because of competition between the different compounds (van Wouwe et al., 2004a). Therefore toxicity estimated from the CALUX bioassay and that from chemical analyses could differ significantly.

The Special Administrative Region of Hong Kong, together with 25 other countries, participated in the 2002–2003 WHO/EURO co-ordinated dioxins and dioxin-like PCBs exposure study (Malisch and van Leeuwen, 2003). Hong Kong submitted pooled milk samples collected from 316 primiparous women and the pooled WHO-TEQ were determined by GC/MS for comparison of between country differences. (Hedley et al., 2006) The major limitation of this approach is that without extensive stratification in the pooling of samples important individual factors such as age cannot be studied without potential confounding from other characteristics of pooled subjects. In order to determine individual demographic and lifestyle characteristics associated with levels of dioxin exposure in Hong Kong, CALUX bioassay was used to determine the TEQ concentration (CALUX-TEQ) of 250 individual milk samples, which were paired samples of those forming eleven of the Hong Kong pools for the WHO exposure study (Nelson et al., 2006). In this paper we present an analysis of the agreement between the CALUX-TEQ results of our south China milk samples in the WHO survey and the corresponding GC/MS results.

2. Materials and methods

2.1. Sample collection

The method for sample collection followed the standard protocol for the 2002–2003 WHO/EURO co-ordinated dioxin exposure study. Two hundred and fifty paired milk samples collected from primiparous mothers from Hong Kong, south China and overseas, who gave birth to a singleton in Hong Kong from December 2001 to September 2002, were analysed by both GC/MS and CALUX bioassay for determination of TEQ (Hedley et al., 2006; Nelson et al., 2006). Each of these 250 milk samples was divided into 10 ml and 20 ml portions. CALUX bioassay was carried out on each 10 ml milk sample and the 20 ml portions were used to form eleven pools. The pools were created with the highest possible internal homogeneity and between-pool diversity achievable in terms of geographic origin defined as Hong Kong, mainland China and Overseas, and dietary (dairy products and seafood) exposures, for the GC/MS analysis (Table 1). This study was approved by the institutional review boards of Hong Kong University, the Chinese University of Hong Kong and the Department of Health, Hong Kong SAR Government. All participants gave written consent before taking part in the study.

2.2. Determination of dioxin and dioxin-like toxicities

GC/MS analyses were carried out by the reference laboratory for WHO/EURO co-ordinated dioxin exposure study, the State Institute for Chemical and Veterinary Analysis of Food, Freiburg, Germany (Malisch and van Leeuwen, 2002) and CALUX bioassay was performed by the BioDetection Systems b.v. (http://www.biodetection-systems.com/) in the Netherlands. For ease of comparison, dioxin levels obtained from the GC/MS analysis were presented as picogram of WHO-TEQ per gram of fat in breast milk (pg WHO-TEQ/g fat) and the dioxin levels estimated

Table 1
Difference (mean, 95% CI) between CALUX-TEQ and 1997-WHO-TEQ and 2005-WHO-TEQ among 11 pooled milk samples

Pool characteristics[a]		Number of subjects	Mean age (years)	CALUX-TEQ (pg/g fat) Mean ± S.D.[b]	1997-WHO-TEQ/(pg/g fat)		2005-WHO-TEQ/(pg/g fat)	
					TEQ_{WHO}[c]	$(TEQ_{CALUX\text{-}WHO})$[d]	TEQ_{WHO}[c]	$(TEQ_{CALUX\text{-}WHO})$[d]
1. Hong Kong	(ever-smokers)	25	26.5	11.8 ± 5.8	10.6	1.3	8.5	3.3
2. Hong Kong	(high dairy product intake)	16	31.0	14.5 ± 4.2	13.7	0.8	11.2	3.3
3. Hong Kong	(high seafood intake)	17	30.5	16.3 ± 6.2	15.2	1.2	12.4	3.9
4. Hong Kong	(low dairy AND low seafood intake)	32	31.4	14.9 ± 5.6	13.4	1.5	10.7	4.2
5. Hong Kong	(high dairy OR high seafood intake)	28	32.6	16.2 ± 6.3	16.7	−0.4	13.0	3.3
6. Mainland China	(low dairy AND low seafood intake)	21	26.6	12.2 ± 5.0	9.0	3.2	7.7	4.5
7. Mainland China	(high dairy OR high seafood intake)	34	28.1	14.3 ± 7.0	11.6	2.7	10.0	4.3
8. China Immigrant	(2–6 years stay in Hong Kong)	22	27.6	14.2 ± 6.1	10.9	3.3	9.2	5.0
9. China Immigrant	(7 years stay in Hong Kong or above)	23	30.1	16.4 ± 4.5	13.5	2.9	11.4	5.0
10. Overseas	(1–10 years of overseas stay)	18	32.1	15.5 ± 4.4	15.0	0.5	12.1	3.4
11. Overseas	(11 years of overseas stay or above)	10	25.7	11.5 ± 6.1	11.3	0.2	8.4	3.1
Overall		246	30.0	14.5 ± 5.8	12.8[e]	1.6 [0.7, 2.4]	10.5[e]	3.9 [3.5, 4.4]

[a] High dairy product intake: 17 highest dairy product intake and dairy product intake >2.5 × seafood intake. (Seafood includes riverine and marine fish and seafood.) High seafood intake: 17 highest seafood intake and seafood intake >2.5 × dairy product intake. Low dairy AND low seafood: <2 kg/mth dairy product intake and <2 kg/mth seafood intake. High dairy OR high seafood: ⩾2 kg/mth dairy product intake OR ⩾ 2 kg/mth seafood intake.
[b] S.D. = Standard deviation.
[c] WHO-TEQ derived from mass concentrations determined by GC/MS and the TEF of the 29 target congeners.
[d] TEQ_{CALUX}-TEQ_{WHO} (pg/g fat).
[e] Mean for WHO-TEQ was weighted by the number of subjects in each pool.

by the CALUX bioassay as picogram CALUX-TEQ per gram of milk fat (pg CALUX-TEQ/g fat).

The mass concentrations of the targeted 7 PCDD, 10 PCDF and 12 PCB congeners of the eleven pools were determined by GC/MS chemical analyses in 2003. The toxicity of 7 PCDD, 10 PCDF and 12 PCB congeners were expressed as TEQ which was the sum of the mass concentrations of the individual dioxin or dioxin-like compounds multiplied by their respective TEFs derived in 1997 (TEF_{1997}) (van den Berg et al., 1998). In 2005, the TEF system was revised (van den Berg et al., 2006) and a set of new TEFs (TEF_{2005}) was also applied to calculate the WHO-TEQ for comparison. To distinguish the WHO-TEQs calculated with TEF_{1997} and TEF_{2005}, they are presented as 1997-WHO-TEQ and 2005-WHO-TEQ respectively.

The CALUX bioassay comprises a genetically modified H4IIE rat hepatoma cell-line, incorporating the firefly luciferase gene coupled to dioxin responsive elements (DREs) as a reporter gene for the presence of dioxins and dioxin-like compounds (Aarts et al., 1996). Fat soluble compounds were extracted from milk by means of shaking with hexane: diethylether (97:3). Fat extracted from the samples was cleaned on an acid silica column (20% and 33% H_2SO_4), topped with sodium sulphate, to remove non-stable compounds. The cleaned extracts were evaporated and redissolved in 25 μl DMSO, after which the CALUX activity was determined on 96-well microtiter plates (24 h exposure; 0.4% DMSO). Dioxins and dioxin-like compounds in the cleaned extracts bind to cytosolic Ah-receptor and the complex was translocated to the nucleus of the cell where it induced the transcription of the recombinant gene. Luciferase was produced and by addition of the substrate luciferin, light emitted. The amount of light being produced was proportional to the amount of ligand–Ah-receptor binding, which was expected to be directly proportional to the toxicity mediated by the Ah-receptor. On each 96-well microtiter plate, a standard TCDD calibration curve was constructed by exposing cells to quantitatively prepared serial dilutions of TCDD dimethylsulphoxide. In this way, the CALUX bioassay reported TEQs benchmarked against TCDD as CALUX-TEQs. CALUX-TEQ was the only product of the CALUX bioassay used this time (DR-CALUX®, dioxin responsive-CALUX). Each milk sample was analysed at multiple dilutions and for each dilution a triplicate assay was performed. Procedure blanks and DMSO blanks were used to control the quality of the analysis. To determine the recovery in the DR-CALUX® bioassay, multiple of analyses of reference samples per series of unknown samples analysed were performed. The results for the reference sample were plotted on a Shewhart Chart. The result of an assay was considered correct if the results were within

the 95% CI (Confidence Intervals) of the Shewhart Chart. If results were outside the 95% CI, the series was considered false and those samples were re-analysed.

2.3. Data analysis

The individual CALUX values for each pool are presented as means and standard deviations (S.D.). The differences between the WHO-TEQ (derived from GC/MS) and CALUX-TEQ were calculated for each pool. For each milk sample the difference between the two estimates was plotted against the mean of the two estimates, to highlight the relationship between the two measures (Bland and Altman, 1986). The second visualisation was made by plotting the WHO-TEQ against CALUX-TEQ for each paired milk sample.

The agreement between the mean CALUX-TEQ and the pooled results by the GC/MS among the eleven pools was then assessed by adopting the approach recommended by Altman and Bland (Altman and Bland, 1983; Bland and Altman, 1986). The difference d between the two measurements was calculated and a 95% confidence interval for d was calculated from the appropriate t distribution. The two measurement methods could be used interchangeably to obtain a single estimate of dioxin activity when the differences d between CALUX-TEQ and WHO-TEQ were considered to be acceptable.

We formulated a simple regression model for WHO-TEQ on CALUX-TEQ, and two multiple regression models to predict WHO-TEQ from CALUX-TEQ, where the first allowed for constant systematic errors depending on the residential characteristics of the mothers and the second additionally allowed for a proportional systematic error by the level of toxicity (van Wouwe et al., 2004b). Multiple regression models were compared based on the adjusted R-square statistics. We first performed the above analyses using 1997-WHO-TEQ and presented the results as our main findings. Then all analyses for agreement were repeated for the 2005-WHO-TEQ for comparison. Data analysis was performed by using the Statistical Package for Social Sciences (SPSS for Windows, version 10.1; SPSS Inc., Chicago, United States) and R version 2.3.1 (R Development Core Team, Vienna, Austria).

3. Results and discussion

3.1. Subject characteristics

The majority (98%) of the 250 mothers (aged 17–42 years old) were Chinese; about half (52%; $n = 130$) were born in Hong Kong, 43% ($n = 107$) were born in mainland China and the remainder in other Asia Pacific countries. Sixty-nine mainland mothers came to Hong Kong prior to delivery and their mean duration of stay in Hong Kong was less than a year. The mean age of donors in each of the eleven pools ranged from 26.5 years to 32.6 years (Table 1). The mean age of babies at the time of sampling the mothers was 4.1–4.6 weeks. Mothers who recently came from mainland China (Pools 6–8) were likely to be younger, less educated and with less household income compared with mothers who resided in Hong Kong for longer periods (Pools 1–5 and 9–10). They were also more likely to practice exclusive breastfeeding. Detailed characteristics of the participants and their allocation to geographic residential categories have been described elsewhere (Hedley et al., 2006).

3.2. GC/MS analyses

The GC/MS analyses generated the mass concentration for each of the 29 target congeners in the 11 pools. (Hedley et al., 2006) All but two (2,3,4,6,7,8-hexachlorodibenzofuran and PCB123) of the congeners were found to have detectable contents in all the 11 pooled milk samples. The analytical difference between upper and lower bound calculations of TEQ values was negligible (<0.1%), therefore upper-bound concentrations were used. In all pools, the 1997-WHO-TEQ was larger than the corresponding 2005-WHO-TEQ and the difference ranged from 1.3 to 3.72 pg/g fat (Table 1).

3.3. CALUX bioassay

Two hundred and thirty-three out of 250 individual milk samples gave quantifiable responses ranging from 3.2 to 33 pg CALUX-TEQ/g fat. The fat content of one sample was reported as missing and the TEQ content could not be determined. An estimated TEQ was allocated to 13 samples (belonging to pools 1, 2, 6, 7, 8 and 11) with responses just below the limit of quantification and a further 3 were classified as undetectable. As a result, CALUX-TEQ values were obtained for a total of 246 milk samples. Further analysis shows that exclusion of the 13 estimated CALUX-TEQ did not affect the study findings (data not shown). Among mothers with the same residential class (Hong Kong, mainland China, China immigrant and Overseas) and/or similar dietary pattern of dioxin containing foods (dairy products and seafoods), an intra-pool range of 12.9–31.8 pg CALUX-TEQ/g fat was observed. The four samples with no values for CALUX-TEQ were from Pool 2 ($n = 1$), Pool 5 (1) and Pool 11 (2). The CALUX-TEQs of the individual milk samples averaged by pools ranged from 11.5 to 16.4 pg/g fat (Table 1).

3.4. Comparison between WHO-TEQ and CALUX-TEQ

Fig. 1a shows the difference between the two estimates of toxicity for each milk sample (CALUX-TEQ and 1997-WHO-TEQ) plotted against the mean of the two estimates. The CALUX-TEQ was typically higher than the 1997-WHO-TEQ, while the degree of discrepancy appears unaffected by the mean value of the two measurements. A greater discrepancy was observed among the four pools comprising mothers from mainland China compared to

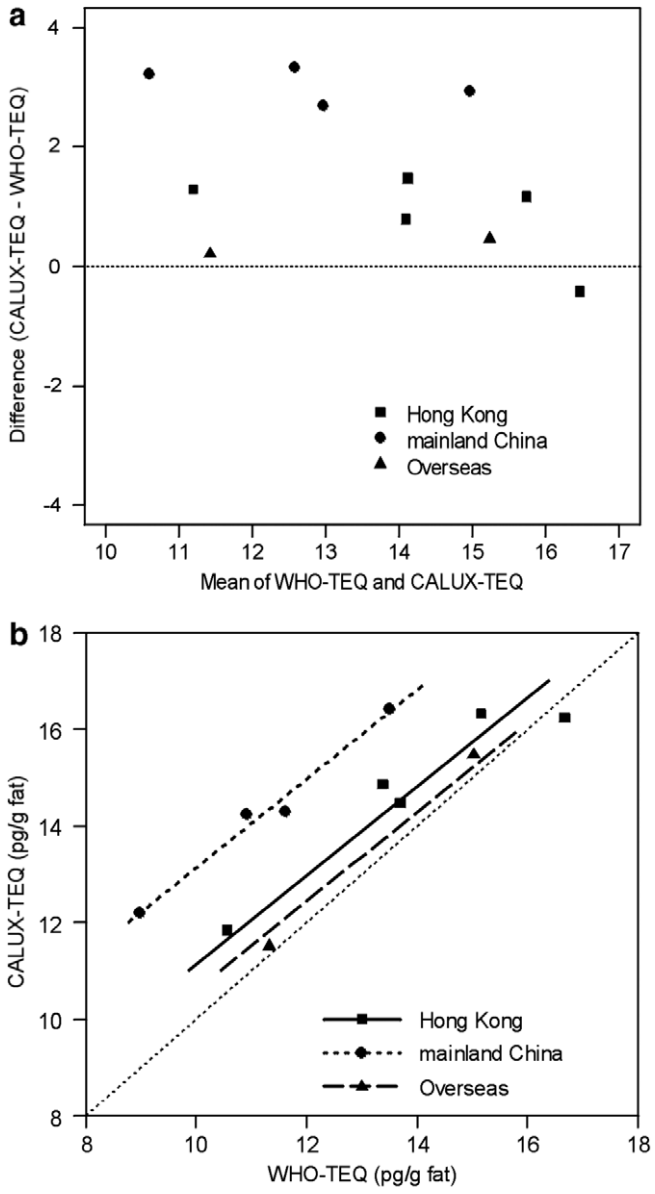

Fig. 1. (a) Difference against mean of CALUX-TEQ and WHO-TEQ in 11 pools and (b) comparison of the mean CALUX-TEQ and the 1997-WHO-TEQ in 11 pools.

the pools of Hong Kong mothers and those who had spent time overseas (Fig. 1a). When the 1997-WHO-TEQ was plotted against average CALUX-TEQ of the corresponding pool, the lowest level of agreement between the two measurements, that is those furthest away from the line of equality, was found in the four pools comprising mothers who came from mainland China (Fig. 1b).

The actual difference between the two TEQ estimates in each pool was calculated (Table 1). Compared with the 1997-WHO-TEQ, a higher CALUX-TEQ was observed in each pool except Pool 5 comprising Hong Kong residents with high intake of seafood and dairy products (Table 1). The overall mean difference between the two TEQs was 1.6 pg/g fat (95% CI: 0.7, 2.4). When the TEF_{2005} was applied, there was a 14–26% decrease in the 2005-WHO-TEQ resulting in a larger the difference between the CALUX- and WHO-TEQs (mean: 3.9 pg/g fat; 95% CI: 3.5–4.4).

A simple linear regression of 1997-WHO-TEQ (y) on CALUX-TEQ (x) obtained the equation $y = -2.6 + 1.08x$, (adj. $R^2 = 0.66$). However given the clear relationship between the residential characteristics and the degree of deviation (Fig. 1a), we investigated two multiple regression models (Table 2). In Model 1, the discrepancy between 1997-WHO-TEQ and CALUX-TEQ was allowed to vary by residential characteristics of the mothers, while Model 2 adjusted for residential characteristics and additionally allowed the difference to vary by the level of CALUX-TEQ i.e. the discrepancy was proportional to the CALUX-TEQ value. The adjusted R-square for Model 2 was higher (adj. $R^2 = 0.92$) than that for Model 1 (adj. $R^2 = 0.81$) indicating that the inclusion of a term for the proportional systematic error is warranted. In a sensitivity analysis, Model 2 fitted better than Model 1 after exclusion of the outlying point for Pool 5, for which the CALUX-TEQ was lower than its 1997-WHO-TEQ. These results suggested that there was a slight ($\beta = 1.1$) proportional systematic error where the CALUX bioassay gave greater overestimation of lower toxicities (Table 2). According to Model 1 and Model 2, a CALUX-estimated toxicity of 13 pg/g fat would overestimate the 1997-WHO-TEQ by approximately 0.9, 3.1 and 0.3 pg/g fat for mothers from Hong Kong, mainland China and Overseas, respectively. In the regression model of 2005-WHO-TEQ on CALUX-TEQ, the regional difference in the discrepancy between the two measurements was narrower and the small proportional systematic error disappeared (data not shown).

3.5. Interpretation of the study findings

The bioanalytical technique in quantification of dioxin levels in humans is increasingly being explored and used in population research, however, the interpretation of the

Table 2
Multiple regression models to predict 1997-WHO-TEQ from CALUX-TEQ

	Model 1[a] β [95% CI]	Model 2[b] β [95% CI]
CALUX-TEQ[c]	1	1.1 [−0.2, 0.3]
Residential characteristics		
Hong Kong	−0.9 [−1.5, −0.3]	−2.2 [−6.0, 1.5]
Mainland China	−3.1 [−3.7, −2.4]	−4.4 [−8.0, −0.7]
Overseas	−0.3 [−1.2, 0.6]	−1.6 [−5.1, 1.9]
R^2	0.85	0.95
Adjusted R^2	0.81	0.92

[a] Allowing for discrepancies related to the residential characteristics of the mothers.
[b] Allowing for discrepancies related to the level of CALUX-TEQ and the residential characteristics of the mothers.
[c] In Model 1, the slope of this regression coefficient was restricted to be 1. In Model 2, we allowed for a proportional systematic error i.e. discrepancies related to the level of CALUX-TEQ.

dioxin toxicity determined by CALUX bioassay is more complex than the WHO-TEQ (Windal et al., 2005). The value of the CALUX bioassay in surveillance work depends on the assessment of the validity of the CALUX-TEQ. There are several reports on the comparison between CALUX results and those from GC/MS on human milk samples (Aarts et al., 1996; Laier et al., 2003; Kayama et al., 2003) and human blood samples (Koppen et al., 2001; Kayama et al., 2002; van Wouwe et al., 2004b; Warner et al., 2005). Most of them assessed the relationship between the two measurements by calculating the correlation coefficients. However high correlation does not necessarily mean the two measures are interchangeable and comparable. Moreover, the typical hypothesis test against a correlation coefficient of 0 is not as useful when the two measures are expected to be correlated (Bland and Altman, 1986). In this study we have adopted another approach suggested by Bland and Altman (1986) and also multiple regression modelling to further investigate the interpretation of the CALUX-TEQ.

In this study, the 95% confidence interval of the difference between the two measurements among the 11 paired samples did not include zero indicating that there was a statistically significant difference between the WHO-TEQ and CALUX-TEQ. An even larger discrepancy was observed when the TEF_{2005} was used to calculate the WHO-TEQ as the change in TEF system resulted in about a 20% (14–26%) decrease in the TEQ, as also suggested by van den Berg et al. (2006). The GC/MS measures the concentrations of a selected number of dioxin and dioxin-like compounds while the CALUX bioassay detects all compounds acting as Ah-receptor agonists so the systematically higher values with bioassays observed in this study were plausible and in line with some previous reports on human samples (Aarts et al., 1996; Laier et al., 2003; van Wouwe et al., 2004b). Illegal electronic waste recycling has been prevalent in certain region of southern China (Puckett et al., 2002) and we believe part of the overestimation of CALUX-TEQ was probably contributed from the presence of compounds possessing Ah-receptor activity (Behnisch et al., 2003), such as polybrominated diphenyl ethers, polybrominated dibenzo-*p*-dioxins and polybrominated dibenzofurans, in the matrix. It has been suggested that the use of lower bound concentrations determined by GC/MS analysis could induce greater discrepancies between the two measurements at lower toxicities (van Wouwe et al., 2004b). However this could not explain the observations in our data because the analytical difference between upper-bound and lower-bound calculations of WHO-TEQ is negligible (<0.1%) and upper-bound concentrations were used.

The deviation of CALUX-TEQ from WHO-TEQ was greater in milk samples from recent (Pools 6 and 7) and earlier (Pools 8 and 9) immigrants from mainland China while the smallest deviation was observed among mothers who had spent some time away from Hong Kong (Pools 10 and 11). The residential characteristics of the mothers significantly improved prediction of WHO-TEQ by CALUX-TEQ in the multiple linear regression model. Although the newly published TEFs resulted in a narrower variation in the size of overestimation, regional differences remained and this effect is evident in regression models on the 2005-WHO-TEQ. These observations indicated that taking into account differences in the residential histories of the mothers, even within the same region, improved the interpretation of the CALUX-TEQ in their breast milk samples and this should be considered in inter-country comparisons of CALUX-TEQ. Regional differences in the overestimation of CALUX-TEQ in paired human serum has been reported among subjects living in urban and rural areas (Koppen et al., 2001), implying that the source of the discrepancy is geographic specific and related to the environmental exposure of the mothers. Milk samples from the mainland China had a higher % PCDF-TEQ of the total PCDD/F-TEQ (44–45%) and a lower % mono-*ortho* PCBs in the total PCB-TEQs (23.4–32.2%) compared to those for Hong Kong (respectively 37–38% and 41.3–49.7%) (Hedley et al., 2006). Such differences in PCDD, PCDF and PCB congener profiles might also contribute to the discrepancy in the bioassay performance. For example, Hoogenboom et al. (2001) has reported an underestimation of TEQ in the bioassay with increasing concentration of mono-*ortho* PCB in eel samples as some of them showed a relatively weak response in the bioassay compared with the WHO-TEF (Gizzi et al., 2005). However since we did not have information on the CALUX-TEQ of individual congeners we could not tell exactly how the congener profiles and difference in REPs and TEFs interacted in the regional differences in the CALUX-TEQ and WHO-TEQ.

There has been no established acceptable range of difference between the CALUX-TEQ and WHO-TEQ. In this study, we demonstrated that the valid measure of difference between the assays varied between subjects from different regions. A judgement on what is an "acceptable difference" (Bland and Altman, 1986) would be based on the absolute magnitude and dispersion of the TEQs, and the size of the difference between the assays. We believe that the size of the differences in the two estimates for Hong Kong and Overseas samples are small and implicitly acceptable. The departure from this pattern in mainland China samples, where the CALUX appeared to over-estimate TEQ compared with GC/MS, clearly points to potentially important issues of interpretation when the assay is used across populations with marked regional differences in social demographic and environmental characteristics. We found a small proportional systematic error (van Wouwe et al., 2004b), which indicated that the relationship between WHO-TEQ and CALUX-TEQ was also dependent on the actual level of toxicity with larger CALUX over-estimation for lower toxicities. This effect was not very strong and it even disappeared when the 2005-WHO-TEQ was used for comparison. The deviation of CALUX-TEQ from WHO-TEQ was more dependent on the geographic characteristics of the mothers.

In our defined population of nursing mothers, using paired milk samples, we showed there was a significant difference between the gold standard WHO-TEQ and the CALUX-TEQ. The CALUX bioassay generally gave an over-estimation of the toxicity (of 0.2–3.3 pg/g fat) compared with the GC/MS method but the degree of over-estimation varied among different geographic origins of the mothers. Since the TEFs and relative potencies for some congeners are different and the congener profiles vary between geographic regions, our findings cannot be generalised to all regions. The comparison between the two measurements is however altered when the new TEF system (van den Berg et al., 2006) is used. In this new system, TEFs for 14 congeners were revised among which 7 were found to be closer to the relative effect potency (REP) used for CALUX bioassay while 7 were more separated. After applying the new TEF, the overestimation by the CALUX-TEQ was found to be two to three times greater and the regional difference in the discrepancy between the two measurements was narrower. More studies are needed to investigate the effect of the new-TEF system on the interpretation of CALUX-TEQ.

There are a few limitations in this study. First, the relatively small sample size within multiple pools limited the power of the multiple regression analyses and did not allow the investigation of other discrepancies between CALUX-TEQ and WHO-TEQ, which may arise because of variations in concentrations of probable agonists. Second, since CALUX-TEQ was not determined on pooled milk samples, we cannot affirm whether the arithmetic mean of individual CALUX-TEQ is exactly equivalent to the CALUX-TEQ of the pooled samples. Third taking the mean of TEQ, in either individual or pooled samples, reduces the dispersion of the data which could over simplify the comparison between WHO-TEQ and CALUX-TEQ and would likely cause us to overestimate the fit (R^2) of the regression models.

In view of the fact that CALUX measurements do not always reflect real dioxin-like toxicity and that its performance may be affected by a number of different factors (Windal et al., 2005), the use of the CALUX bioassay must be combined with appropriate and adequate extraction and clean-up procedures to prevent non-dioxin responses (van Overmeire et al., 2001) as this is crucial to ensure compatibility (Harrison and Eduljee, 1999) and interpretation (van Wouwe et al., 2004a) of bioanalytical detection methods. Knowledge of substances with potential CALUX assay activities in any specific sample matrix, such as endogenous activators or other sources of pollutants, is also needed (Ziccardi et al., 2000). Validation with toxicity determined by chemical analysis on paired samples would improve interpretation of the bioassay results in population surveys.

Acknowledgments

This work was supported by the Hong Kong Environment Conservation Fund, Grant 8/2000.

We thank all the mothers who donated milk samples and Dr. A. Chiu, Dr. S. Leung and all the nursing staff in Maternity and Child Health Centres for their help in recruitment of the mothers and sample collection.

We are very grateful to Dr. K. Kypke for performing and monitoring the GC/MS chemical analysis and Prof. B. Brouwer and Dr. H. Besselink for performing the CALUX bioassay. We also thank Ms D. Wong and Ms H. Choi for collection of milk samples and demographic data and Dr. C.M. Wong and Dr. T.Q. Thach for statistical advice. We thank Mr. S.T. Kwok for technical help in the laboratory and Ms M. Chi for secretarial support.

References

Aarts, J.M.M.J.G., Cenijn, P.H., Blankvoort, B.M.G., Murk, A.J., Brouwer, A., 1996. Application of the chemical-activated luciferase gene expression (CALUX) bioassay for quantification of dioxin-like compounds in small samples of human milk and blood plasma. Organohal. Compd. 27, 285–290.

Altman, D.G., Bland, J.M., 1983. Measurement in medicine – the analysis of method comparison studies. Statistician 32, 307–317.

Behnisch, P.A., Allen, R., Anderson, J., Brouwer, A., Brown, D.J., Campbell, C., Goeyens, L., Harrison, R.O., Hoogenboom, R., Van Overmeire, I., Traag, W., Malisch, R., 2001a. Harmonised quality criteria for chemical and bioassays analyses of PCDDs/PCDFs in feed and food. Part 2: general considerations, bioassay methods. Organohal. Compd. 50, 59–63.

Behnisch, P.A., Hosoe, K., Sakai, S.I., 2001b. Bioanalytical screening methods for dioxins and dioxin-like compounds – a review of bioassay/biomarker technology. Environ. Int. 27, 413–439.

Behnisch, P.A., Hosoe, K., Sakai, S.I., 2003. Brominated dioxin-like compounds: in vitro assessment in comparison to classical dioxin-like compounds and other polyaromatic compounds. Environ. Int. 29, 861–877.

Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1, 307–310.

Butler, R.A., Kelley, M.L., Olberding, K.E., Gardner, G.R., van Beneden, R.J., 2004. Aryl hydrocarbon receptor (AhR)-independent effects of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) on softshell clam (*Mya arenaria*) reproductive tissue. Comp. Biochem. Physiol. C. 138, 375–381.

European Commission. 2002a, Commission directive 2002/69/EC of 26 July 2002. Laying down the sampling methods and the methods of analysis for the official control of dioxins and the determination of dioxin-like PCBs in foodstuffs. Official Journal L 209.

European Commission. 2002b, Commission directive 2002/70/EC of 26 July 2002. Establishing requirements for the determination of levels of dioxins and dioxin-like PCBs in feedingstuffs. European Union. Official Journal L 209.

Garrison, P.M., Tullis, K., Aarts, J.M., Brouwer, A., Giesy, J.P., Denison, M.S., 1996. Species-specific recombinant cell lines as bioassay systems for the detection of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin-like chemicals. Fundam. Appl. Toxicol. 30, 194–203.

Giesy, J.P., Hilscherova, K., Jones, P.D., Kannan, K., Machala, M., 2002. Cell bioassays for detection of aryl hydrocarbon (AhR) and estrogen receptor (ER) mediated activity in environmental samples. Mar. Pollut. Bull. 45, 3–16.

Gizzi, G., Hoogenboom, L.A.P., Von Holst, C., Rose, M., Anklam, E., 2005. Determination of dioxins (PCDDs/PCDFs) and PCBs in food and feed using the DR-CALUX bioassay: results of an international validation study. Food Addit. Contam. 22, 472–481.

Harrison, R.O., Eduljee, G.H., 1999. Immunochemical analysis for dioxins – progress and prospects. Sci. Total Environ. 239, 1–18.

Hedley, A.J., Wong, T.W., Hui, L.L., Malisch, R., Nelson, E.A.S., 2006. Breast milk dioxins in Hong Kong and Pearl River Delta. Environ. Health Perspect. 114, 202–208.

Hoogenboom, L.A.P., Traag, W.A., Hoogerbrugge, R., Baumann, B., De Vreies, J., 2001. Control analysis of eel using CALUX and GC/MS. Organohal. Compd. 54, 23–27.

JECFA, 2002. WHO Food Additive Series: 48 Safety Evaluation of Certain Food Additives and Contaminants. WHO, Geneva.

Kayama, F., Horiguchi, H., Oguma, E., Fujino, J., Yabushita, H., Brown, D., Clark, G., 2002. Regional difference of blood dioxin and organochlorine pesticides concentrations of Japanese female farmers. Application of CALUX assay for epidemiological study. Organohal. Compd. 55, 275–278.

Kayama, F., Horiguchi, H., Jin, Y., Kunisue, T., Tanabe, S., Fukatsu, H., Yamazaki, N., Nakamura, M., Yabushita, H., Fujino, J., Clark, G., 2003. Dioxin and organochlorine pesticide concentrations of breast milk in China and Japan. Organohal. Compd. 64, 207–210.

Koppen, G., Covaci, A., van Cleuvenbergen, R., Schepens, P., Winneke, G., Nelen, V., Schoeters, G., 2001. Comparison of CALUX-TEQ values with PCB and PCDD/F measurements in human serum of the Flanders Environmental and Health Study (FLEHS). Toxicol. Lett. 123, 59–67.

Laier, P., Cederberg, T., Larsen, J.C., Vinggaard, A.M., 2003. Applicability of the CALUX bioassay for screening of dioxin levels in human milk samples. Food Addit. Contam. 20, 583–595.

Malisch, R., van Leeuwen, F.X.R., 2002. Third round of WHO-coordinated exposure study: analysis of PCDDs, PCDFs and PCBs in human milk. Organohal. Compd. 56, 317–320.

Malisch, R., van Leeuwen, F.X.R., 2003. Results of the WHO-coordinated exposure study on the levels of PCBs, PCDDs and PCDFs in human milk. Organohal. Compd. 64, 140–143.

Nelson, E.A.S., Hui, L.L., Wong, T.W., Hedley, A.J., 2006. Factors associated with dioxin-like activity in human breast milk in Hong Kong. Environ. Sci. Technol. 40, 1432–1438.

Puckett, J., Byster, L., Westervelt, S., Gutierrez, R., Davis, S., Hussain, A., 2002. Exporting Harm: The High-Tech Trashing of Asia. The Basel Action Network, Silicon Valley Toxics Coalition, Seattle, WA.

Safe, S., 1993. Development of bioassays and approaches for the risk assessment of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin and related compounds. Environ. Health Perspect. 101 (Suppl. 3), 317–325.

Seidel, S.D., Li, V., Winter, G.M., Rogers, W.J., Martinez, E.I., Denison, M.S., 2000. Ah receptor-based chemical screening bioassays: application and limitations for the detection of Ah receptor agonists. Toxicol. Sci. 55, 107–115.

van den Berg, M., Birnbaum, L., Bosveld, A.T., Brunstrom, B., Cook, P., Feeley, M., Giesy, J.P., Hanberg, A., Hasegawa, R., Kennedy, S.W., Kubiak, T., Larsen, J.C., van Leeuwen, F.X., Liem, A.K., Nolt, C., Peterson, R.E., Poellinger, L., Safe, S., Schrenk, D., Tillitt, D., Tysklind, M., Younes, M., Waern, F., Zacharewski, T., 1998. Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. Environ. Health Perspect. 106, 775–792.

van den Berg, M., Birnbaum, L.S., Denison, M., De Vito, M., Farland, W., Feeley, M., Fiedler, H., Hakansson, H., Hanberg, A., Haws, L., Rose, M., Safe, S., Schrenk, D., Tohyama, C., Tritscher, A., Tuomisto, J., Tysklind, M., Walker, N., Peterson, R.E., 2006. The 2005 World Health Organization re-evaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds. Toxicol. Sci. 93, 223–241.

van Overmeire, I., Clark, G.C., Brown, D.J., Chu, M.D., Marcus Cooke, W., Denison, M.S., Baeyens, W., Srebrnik, S., Goeyens, L., 2001. Trace contamination with dioxin-like chemicals: evaluation of bioassay-based TEQ determination for hazard assessment and regulatory responses. Environ. Sci. Policy 4, 345–357.

van Wouwe, N., Windal, I., Vanderperren, H., Eppe, G., Xhrouet, C., De Pauw, E., Goeyens, L., Baeyens, W., 2004a. Importance of clean-up for comparison of TEQ-values obtained by CALUX and chemo-analysis. Talanta 63, 1269–1272.

van Wouwe, N., Windal, I., Vanderperren, H., Eppe, G., Xhrouet, C., Massart, A.C., Debacker, N., Sasse, A., Baeyens, W., De Pauw, E., 2004b. Validation of the CALUX bioassay for PCDD/F analyses in human blood plasma and comparison with GC–HRMS. Talanta 63, 1157–1167.

Warner, M., Eskenazi, B., Patterson, D.G., Clark, G., Turner, W.E., Bonsignore, L., Mocarelli, P., Gerthoux, P.M., 2005. Dioxin-Like TEQ of women from the Seveso, Italy area by ID-HRGC/HRMS and CALUX. J. Expo. Anal. Environ. Epidemiol. 15, 310–318.

Windal, I., Denison, M.S., Birnbaum, L.S., Van Wouwe, N., Baeyens, W., Goeyens, L., 2005. Chemically activated luciferase gene expression (CALUX) cell bioassay analysis for the estimation of dioxin-like activity: critical parameters of the CALUX procedure that impact assay results. Environ. Sci. Technol. 39, 7357–7364.

Ziccardi, M.H., Gardner, I.A., Denison, M.S., 2000. Development and modification of a recombinant cell bioassay to directly detect halogenated and polycyclic aromatic hydrocarbons in serum. Toxicol. Sci. 54, 183–193.