

The experimental design of postmortem studies: the effect size and statistical power

Joris Meurs¹ 

Accepted: 24 June 2016 / Published online: 13 July 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Purpose The aim is of this study was to show the poor statistical power of postmortem studies. Further, this study aimed to find an estimate of the effect size for postmortem studies in order to show the importance of this parameter. This can be an aid in performing power analysis to determine a minimal sample size.

Methods GPower was used to perform calculations on sample size, effect size, and statistical power. The minimal significance (α) and statistical power ($1 - \beta$) were set at 0.05 and 0.80 respectively. Calculations were performed for two groups (Student's *t*-distribution) and multiple groups (one-way ANOVA; *F*-distribution).

Results In this study, an average effect size of 0.46 was found ($n = 22$; $SD = 0.30$). Using this value to calculate the statistical power of another group of postmortem studies ($n = 5$) revealed that the average statistical power of these studies was poor ($1 - \beta < 0.80$).

Conclusion The probability of a type-II error in post-mortem studies is considerable. In order to enhance statistical power of postmortem studies, power analysis should be performed in which the effect size found in this study can be used as a guideline.

Keywords Postmortem research · Sample size · Experimental design · Significance · Power · Effect size

Introduction

Prior to conducting research, several considerations have to be made. For example, the required sample size has to be determined [1]. Commonly, this is done by performing a so-called power analysis [1, 2]. In a power analysis, the sample size is calculated by using four parameters: significance (α), statistical power ($1 - \beta$), variance (σ^2), and effect size (d) [1, 3]. A description and the effect on the sample size of each of these parameters is shown in Table 1. In order to emphasize the effect of α and $1 - \beta$, the confusion matrix is shown in Fig. 1. Despite α and $1 - \beta$ being mostly straightforward values, determining σ^2 and d is rather difficult [1]. In case two independent means are present, Cohen set values of d at 0.20, 0.50, and 0.80 which represent a small, medium, or large effect size respectively [1]. The effect sizes in case multiple means (multiple groups) are present have been set at 0.10, 0.25, and 0.40, which represent a small, medium, or large effect size respectively. According to Cohen, his set medium value for d represents “an effect likely to be visible to the naked eye” [1]. For instance, this can be a change in decomposition stage of a cadaver. In quantitative research this visible effect could be, for example, a significant change in concentration of a certain analyte in a post-mortem sample. Nevertheless, for inexperienced individuals it still remains unclear what the actual meaning of d is. The effect size is defined as the absolute difference between two independent means and the within-sample standard deviation [1, 4]. In other words, how much does a certain situation (e.g., a qualitative or quantitative experiment) differ from reality? Moreover, for calculating d values the independent means (μ_a ; μ_b) and the within-sample standard deviation (σ) have to be estimated [1]. Hence, the resulting d will be a rather subjective value. To solve this

✉ Joris Meurs
j.meurs@student.vu.nl

¹ Department of BioAnalytical Chemistry, VU University Amsterdam, De Boelelaan, 1081 HV Amsterdam, The Netherlands

Table 1 Description and effect of parameters on sample size

Parameter	Description	Effect on sample size
Alpha (α)	The probability of falsely rejecting the null hypothesis (H_0) (i.e., false positive result or type I error) ^a [5]	The lower α , the higher the sample size
Beta (β)	The probability of falsely accepting the null hypothesis (H_0) (i.e., false negative result or type II error) ^a [5]	The lower β , the higher the sample size
Power ($1 - \beta$)	The probability of correctly rejecting the null hypothesis (H_0) ^a [5]	The higher $1 - \beta$, the higher the sample size
Effect size (d or f)	Degree of deviation of an experimental situation compared to an actual situation (i.e., how much does an experiment deviate from reality) [5]	The higher d or f , the lower the sample size
Variance (σ^2)	Expression of the spreading of data around a mean value [5]	The higher σ^2 , the higher the sample size
Noncentrality parameter (λ)	Degree of deviation from the original distribution [15]	$\lambda = 0$: original distribution $\lambda > 0$: increasing noncentrality

^a See Fig. 1 for a graphical explanation

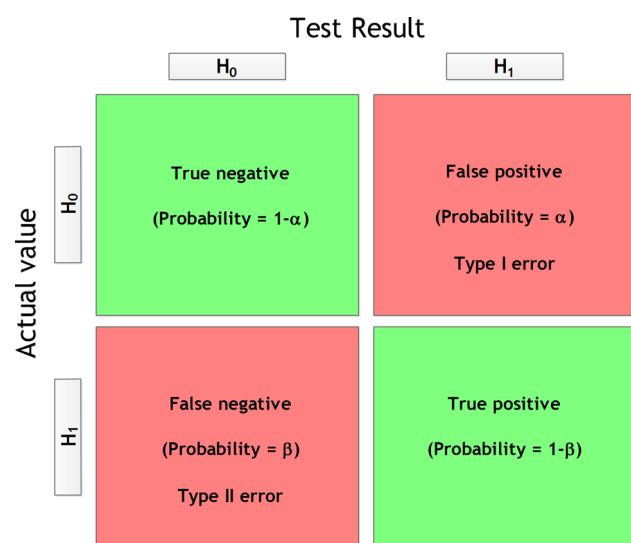


Fig. 1 The confusion matrix of accepting or rejecting the null hypothesis (H_0) or the alternative hypothesis (H_1)

problem, a pilot study can be performed and a sample standard deviation can be used for calculating the effect size [3, 4]. However, pilot studies lack statistical power [5]. Hence, performing a pilot study is not desirable.

It is observed that in postmortem research the sample size is variable. For instance, the sample size can be as low as nine [6] or as high as 57,903 [7]. Low availability of samples or legal restrictions can be a reason for small sample sizes. Although, parameters like the statistical power should still be taken into account despite these limitations. No discussion on the sample size used or the statistical power reached is seen in most publications. Hence, the probability is of false-negative results cannot be

derived from the data that is shown [4]. Therefore, the aim of this paper is to show how a minimal sample size can be estimated without a priori knowledge on the standard deviation to ensure sufficient statistical power. Furthermore, the poor statistical power of postmortem studies will be shown.

Calculation of the sample size in general cases

Two independent means (Student's t test)

To calculate the sample size (n) in order to compare two independent means, Eq. 1 has to be solved [4].

$$n = \left((z\alpha/2 + z\beta) \frac{\sigma}{\delta} \right)^2 \quad (1)$$

where, z is the corresponding z score for values of α and β and δ is defined as the absolute difference between the experimental mean (μ_a) and the control mean (μ_b) (Eq. 2).

$$\delta = |\mu_a - \mu_b| \quad (2)$$

To calculate the z score, values for α were set at 0.05 and 0.01 respectively. Likewise, values for β were set at 0.20, 0.10, and 0.05 respectively. All obtained values are shown below in matrix Z . Column 1 and 2 contain the values for significance levels of 0.05 and 0.01. Values for β decrease going down the rows.

$$Z = \begin{bmatrix} \alpha = 0.05; \beta = 0.20 & \alpha = 0.01; \beta = 0.20 \\ \alpha = 0.05; \beta = 0.10 & \alpha = 0.01; \beta = 0.10 \\ \alpha = 0.05; \beta = 0.05 & \alpha = 0.01; \beta = 0.05 \end{bmatrix} = \begin{bmatrix} 7.849 & 11.6790 \\ 10.5074 & 14.8794 \\ 12.9947 & 17.8142 \end{bmatrix} \quad (3)$$

According to Cohen, the effect size is considered as small, medium, or large at values of 0.20, 0.50, and 0.80 respectively [1]. Since σ/δ is inversely related to the effect size, σ/δ -values of 5, 2, and 1.25 can be considered as large, medium, and small respectively. Therefore, values for the ratio σ^2/δ^2 were set from 0 to 5. With these values, the corresponding sample size (n) was calculated (Fig. 2). To obtain a reasonable estimate for the minimal sample size, for all combinations of α and β the sample size was calculated at the maximum ratio of σ^2/δ^2 . These values are shown in Table 2 and Fig. 3.

Multiple means (ANOVA)

In case of multiple means, the sample size should be determined by using ANOVA. The effect size (f) is then expressed as follows (Eq. 4) [1, 8]:

$$f = \frac{\sigma_m}{\sigma} \tag{4}$$

Accordingly, the total sample size is calculated by using Eq. 5, in which N is the total sample size and λ is the noncentrality parameter [9, 10]. This noncentrality parameter is about 1.5 for $\alpha = 0.01$ when $\beta = 0.20$ and about 1 for $\alpha = 0.05$ when $\beta = 0.20$ [10].

$$N = \frac{\lambda}{f^2} \tag{5}$$

For the one-way ANOVA model, Cohen’s values of 0.10, 0.25, and 0.40 were used to calculate the minimal

Table 2 Overview of sample size in case of two independent means (two groups) at common values of α and β at high value of σ^2/δ^2

α	β	Power ($1 - \beta$)	Sample size (n) ^a
0.05	0.20	0.80	39
0.05	0.10	0.90	52
0.05	0.05	0.95	65
0.01	0.20	0.80	58
0.01	0.10	0.90	74
0.01	0.05	0.95	89

Sample size calculated for equal group sizes

^a Values for n are rounded to the nearest integer

sample size at significance levels of 0.05 and 0.01 respectively. These results are shown in Table 3 and Fig. 4.

Statistical power and effect size of postmortem studies

In order to show the poor statistical power of postmortem studies, a number of studies were selected for post hoc testing on the sample size in order to determine the achieved power. For calculations GPower was used [8]. First, the effect size for a number of postmortem studies ($n = 22$) was calculated. This data is shown in Table 4. Significance level and statistical power were set at 0.05 and 0.80 respectively. A mean effect size of 0.46 (SD = 0.30) was obtained.

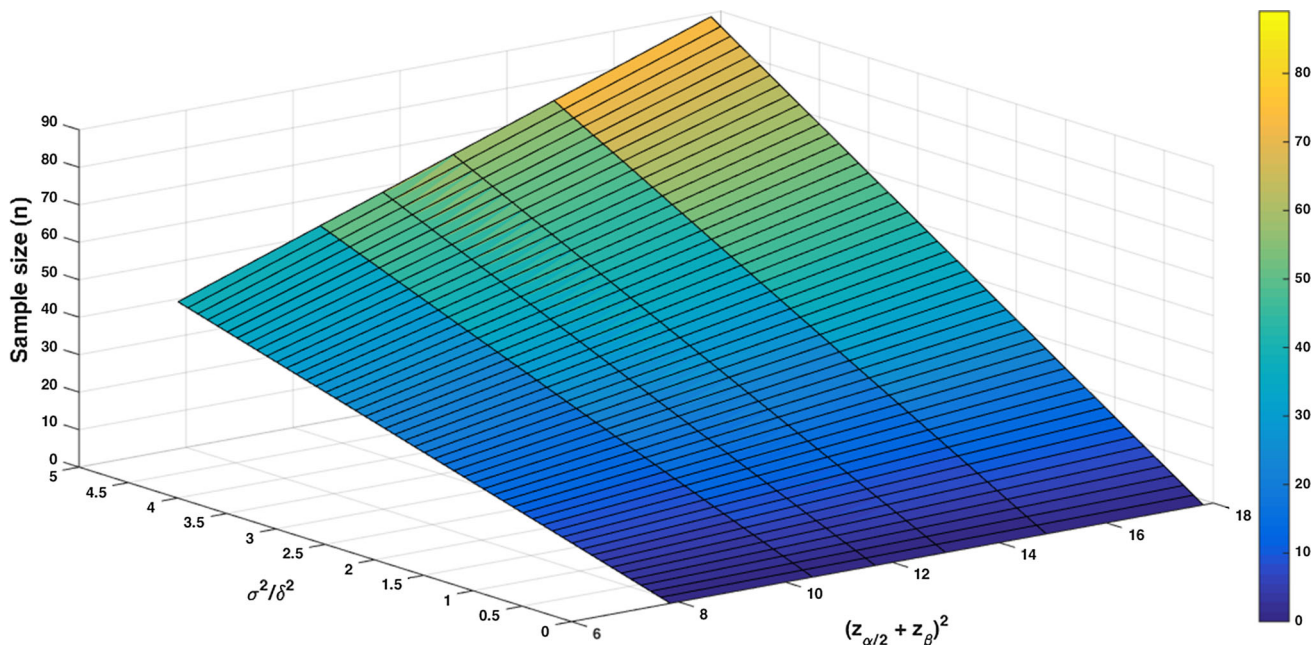


Fig. 2 Influence of $(z_\alpha + z_\beta)^2$ and σ^2/δ^2 on the sample size

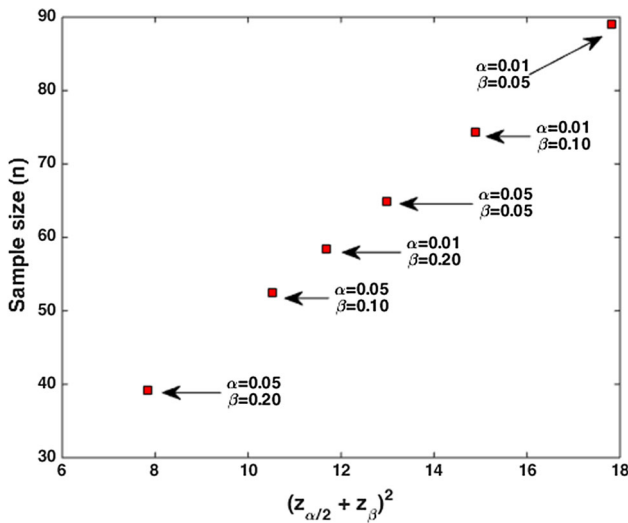


Fig. 3 Sample size for different values of α and β at maximum σ^2/δ^2

This effect size was used to calculate the achieved statistical power of another group of postmortem studies ($n = 5$). A priori, the significance was set at 0.05. The results are shown in Table 5. Only for the studies of Mao et al. [11] and Laiho and Penttilä [12] was the achieved statistical power sufficient (i.e., a value greater than 0.80). In all other cases, the statistical power was less than 0.80, which means there is a reasonable probability of a type-II error. Despite these low power values, the risk of false-negative results are not discussed. An example of a false-negative result is that no significant difference is found in concentration while in fact there is a significant difference. In other words, the null hypothesis (H_0) has been falsely rejected.

Table 3 Overview of sample size in case of multiple means (multiple groups) at common values of α and f ($\beta = 0.20$)

α	f	$N(k = 3)$	$N(k = 4)$	$N(k = 5)$	$N(k = 8)$	$N(k = 10)$
0.01	0.10	1395	1552	1685	1992	2160
0.01	0.25	228	256	275	328	360
0.01	0.40	93	104	115	136	150
0.05	0.10	969	1096	1200	1448	1580
0.05	0.25	159	180	200	240	260
0.05	0.40	66	76	80	104	110

k , group size; values are calculated in GPower [8]

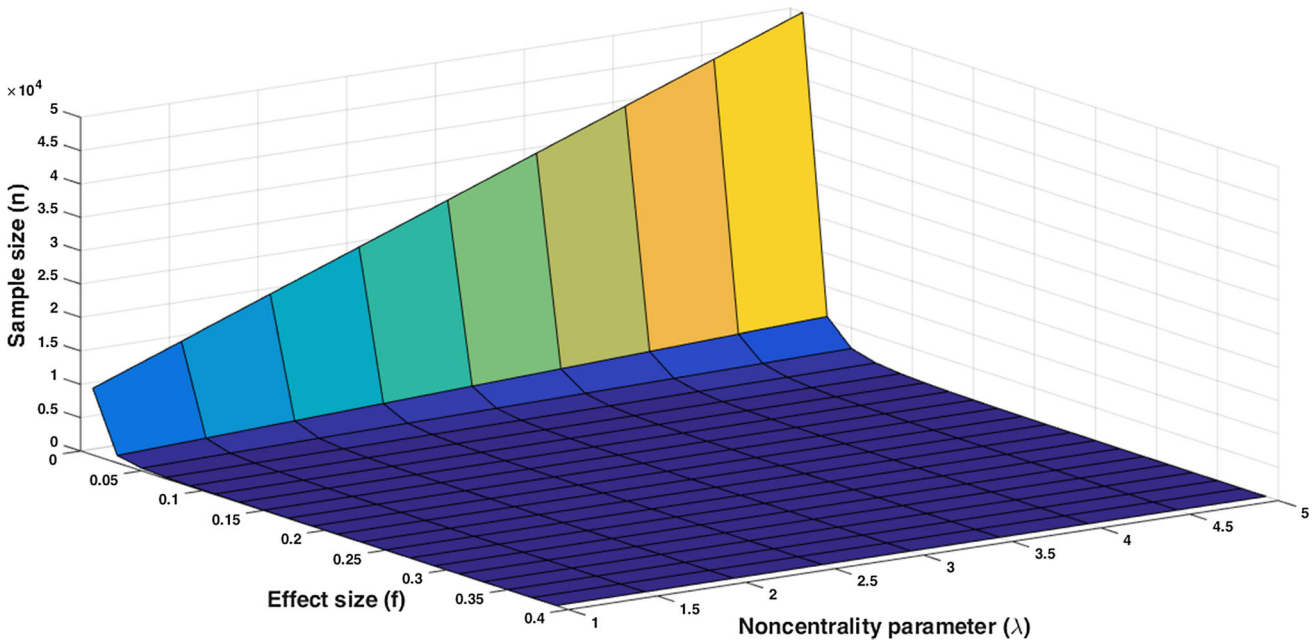


Fig. 4 Influence of f and λ on the sample size

Table 4 Effect size calculation for a number of postmortem studies

References	Sample size (<i>n</i>)	Number of groups	Effect size
Rognum et al. [16]	87	4	0.36
Sato et al. [17]	18	6	1.05
Singh et al. [18]	474	9	0.18
Singh et al. [19]	1026	15	0.13
Wehnet et al. [20]	128	45	0.55
Mihailovic et al. [21]	320	10	0.22
Lemaire et al. [22]	30	4	0.65
Laruelle et al. [23]	34	4	0.60
Pelander et al. [24]	50	2	0.40
Vujanić et al. [25]	540	6	0.15
Krap et al. [26]	10	2	1.01
Li et al. [27]	283	4	0.20
Zhu et al. [28]	405	5	0.17
Koopmanschap et al. [29]	117	3	0.29
Zhu et al. [30]	234	4	0.22
Huang et al. [31]	90	10	0.43
Zheng et al. [32]	111	37	0.56
Li et al. [33]	12	2	0.90
Rognum et al. [34]	32	3	0.58
Maeda et al. [35]	140	4	0.28
Zhu et al. [36]	409	15	0.21
Frere et al. [37]	10	2	1.01
	207.3 ± 246.5	8.9 ± 11.1	0.46 ± 0.30

$\alpha = 0.05$; $\beta = 0.20$; * $p < 0.05$; values are calculated in GPower [8]

Table 5 Post hoc testing performed on a number of postmortem studies ($f = 0.46$)

References	Sample size (<i>n</i>)	Number of groups	Achieved power ($1 - \beta$)
Mao et al. [11]	160	2	0.99
Moriya and Hashimoto [38]	6	2	0.14
Mao et al. [39]	48	6	0.62
Querido and Pillay [40]	36	6	0.46
Laiho and Penttillä [12] ^a	116	8	0.96
	73.2 ± 63.0	4.80 ± 2.68	0.63 ± 0.36

Achieved power was calculated using GPower [8]. Post hoc testing was performed using a one-way ANOVA model with fixed effects

^a Groups were not divided into equal numbers

Discussion and conclusion

Power analysis can be a useful tool in determining the sample size needed for qualitative and quantitative post-mortem experiments. Examples of postmortem qualitative and quantitative research are determining the degree of decomposition [13] and measuring postmortem vitreous potassium [14]. However, in order to calculate the sample size, values have to be set subjectively.

That can be a cause of choosing a random sample size in postmortem research. Sample size determination and

achieved statistical power are rarely discussed in post-mortem studies. However, it is important to discuss these parameters in order to establish the reliability of the obtained results.

This study is the first to demonstrate that postmortem studies lack statistical power. In order to achieve sufficient power, Tables 2 and 3 can be used for obtaining a minimal sample size for common values of significance and statistical power. However, it should always be checked a posteriori if the set levels of power and significance are achieved by performing a post hoc test. Nevertheless,

Tables 2 and 3 can serve as a useful tool in estimating a minimal sample size that would provide sufficient statistical power for postmortem studies.

Additionally, for the first time an estimate of the effect size ($f = 0.46$; $SD = 0.30$) has been shown for post-mortem studies. Besides Tables 2 and 3, this number can be used as an estimate for the effect size in power analysis.

Key Points

1. An effect size has been estimated for postmortem studies.
2. The statistical power of postmortem studies is poor.
3. Power analysis should be performed in order to enhance statistical power of postmortem studies.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Cohen J. A power primer. *Psychol Bull.* 1992;112:155.
2. Dupont WD, Plummer WD. Power and sample size calculations: a review and computer program. *Control Clin Trials.* 1990;11:116–28.
3. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Control Clin Trials.* 1981;2:93–113.
4. Massart DL, Vandeginste BG, Buydens L, Lewi P, Smeyers-Verbeke J. Handbook of chemometrics and qualimetrics: part A. New York: Elsevier Science Inc.; 1997.
5. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: basic principles and common pitfalls. *Nephrol Dial Transpl.* 2010;25:1388–93.
6. Sabucedo AJ, Furton KG. Estimation of postmortem interval using the protein marker cardiac Troponin I. *Forensic Sci Int.* 2003;134:11–6.
7. Launiainen T, Ojanperä I. Drug concentrations in postmortem femoral blood compared with therapeutic concentrations in plasma. *Drug Test Anal.* 2014;6:308–16.
8. Faul F, Erdfelder E, Lang AG, Buchner A. G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 2007;39:175–91.
9. Erdfelder E, Faul F, Buchner A. GPOWER: a general power analysis program. *Behav Res Methods Instrum.* 1996;28:1–11.
10. Pearson ES, Hartley HO. Charts of the power function for analysis of variance tests, derived from the non-central F-distribution. *Biometrika.* 1951;38:112–30.
11. Mao S, Fu G, Seese RR, Wang ZY. Estimation of PMI depends on the changes in ATP and its degradation products. *Leg Med.* 2013;15:235–8.
12. Penttilä A, Laiho K. Autolytic changes in blood cells of human cadavers. II. Morphological studies. *Forensic Sci Int.* 1981;17:121–32.
13. Goff ML. Early postmortem changes and stages of decomposition in exposed cadavers. *Exp Appl Acarol.* 2009;49:21–36.
14. Munoz Barus JI, Suarez-Penaranda J, Otero XL, Rodriguez-Calvo MS, Costas E, Miguens X, et al. Improved estimation of postmortem interval based on differential behaviour of vitreous potassium and hypoxanthine in death by hanging. *Forensic Sci Int.* 2002;125:67–74.
15. Scheffe H. The analysis of variance. New York: Wiley; 1999.
16. Rognum TO, Hauge S, Øyasaeter S, Saugstad OD. A new biochemical method for estimation of postmortem time. *Forensic Sci Int.* 1991;51:139–46.
17. Sato T, Zaitu K, Tsuboi K, Nomura M, Kusano M, Shima N, et al. A preliminary study on postmortem interval estimation of suffocated rats by GC-MS/MS-based plasma metabolic profiling. *Anal Bioanal Chem.* 2015;407:3659–65.
18. Singh D, Prashad R, Parkash C, Sharma SK, Pandey AN. Double logarithmic, linear relationship between plasma chloride concentration and time since death in humans in Chandigarh Zone of North-West India. *Leg Med.* 2003;5:49–54.
19. Singh D, Prashad R, Sharma SK, Pandey AN. Double logarithmic, linear relationship between postmortem vitreous sodium/potassium electrolytes concentration ratio and time since death in subjects of Chandigarh zone of northwest India. *J Indian Acad Forensic Med.* 2005;27:159–65.
20. Wehner F, Wehner HD, Schieffer MC, Subke J. Delimitation of the time of death by immunohistochemical detection of insulin in pancreatic β -cells. *Forensic Sci Int.* 1999;105:161–9.
21. Mihailovic Z, Atanasijevic T, Popovic V, Milosevic MB, Sperhake JP. Estimation of the postmortem interval by analyzing potassium in the vitreous humor: Could repetitive sampling enhance accuracy? *Am J Forensic Med Pathol.* 2012;33:400–3.
22. Lemaire E, Schmidt C, Charlier C, Denooz R, Boxho P. Popliteal vein blood sampling and the postmortem redistribution of Diazepam, Methadone, and Morphine. *J Forensic Sci.* 2016. doi:10.1111/1556-4029.13061.
23. Laruelle M, Abi-Dargham A, Casanova MF, Toti R, Weinberger DR, Kleinman JE. Selective abnormalities of prefrontal serotonergic receptors in schizophrenia: a postmortem study. *Arch Gen Psychiatry.* 1993;50:810–8.
24. Pelander A, Ristimaa J, Ojanperä I. Vitreous humor as an alternative matrix for comprehensive drug screening in postmortem toxicology by liquid chromatography-time-of-flight mass spectrometry. *J Anal Toxicol.* 2010;34:312–8.
25. Vujanić G, Cartlidge P, Stewart J, Dawson A. Perinatal and infant postmortem examinations: How well are we doing? *J Clin Pathol.* 1995;48:998–1001.
26. Krap T, Meurs J, Boertjes J, Duijst W. Technical note: unsafe rectal temperature measurements due to delayed warming of the thermocouple by using a condom. An issue concerning the estimation of the postmortem interval by using Henßge's nomogram. *Int J Leg Med.* 2016;130:447–56.
27. Li DR, Zhu BL, Ishikawa T, Zhao D, Michiue T, Maeda H. Postmortem serum protein S100B levels with regard to the cause of death involving brain damage in medicolegal autopsy cases. *Leg Med.* 2006;8:71–7.
28. Zhu B-L, Ishikawa T, Michiue T, Li DR, Zhao D, Oritani S, et al. Postmortem cardiac troponin T levels in the blood and pericardial fluid. Part 1. Analysis with special regard to traumatic causes of death. *Leg Med.* 2006;8:86–93.
29. Koopmanschap DH, Bayat AR, Kubat B, Bakker HM, Prokorp MW, Klein WM. The radiodensity of cerebrospinal fluid and vitreous humor as indicator of the time since death. *Forensic Sci Med Pathol.* 2016. doi:10.1007/s12024-016-9778-9.
30. Zhu BL, Ishikawa T, Michiue T, Li DR, Zhao D, Bessho Y, et al. Postmortem cardiac troponin I and creatine kinase MB levels in

- the blood and pericardial fluid as markers of myocardial damage in medicolegal autopsy. *Leg Med.* 2007;9:241–50.
31. Huang P, Ke Y, Lu Q, Xin B, Fan S, Yang G, et al. Analysis of postmortem metabolic changes in rat kidney cortex using Fourier transform infrared spectroscopy. *Spectroscopy.* 2008;22:21–31.
 32. Zheng J, Li X, Shan D, Zhang H, Guan D. DNA degradation within mouse brain and dental pulp cells 72 hours postmortem. *Neural Regen Res.* 2012;7:290.
 33. Li X, Greenwood AF, Powers R, Jope RS. Effects of postmortem interval, age, and Alzheimer's disease on G-proteins in human brain. *Neurobiol Aging.* 1996;17:115–22.
 34. Rognum TO, Saugstad OD, Øyasæter S, Olaisen B. Elevated levels of hypoxanthine in vitreous humor indicate prolonged cerebral hypoxia in victims of sudden infant death syndrome. *Pediatrics.* 1988;82:615–8.
 35. Maeda H, Zhu B-L, Ishikawa T, Oritani S, Michiue T, Li DR, et al. Evaluation of postmortem ethanol concentrations in pericardial fluid and bone marrow aspirate. *Forensic Sci Int.* 2006;161:141–3.
 36. Zhu BL, Ishikawa T, Michiue T, Li DR, Zhao D, Quan L, et al. Evaluation of postmortem urea nitrogen, creatinine and uric acid levels in pericardial fluid in forensic autopsy. *Leg Med.* 2005;7:287–92.
 37. Frere B, Suchaud F, Bernier G, Cottin F, Vincent B, Dourel L, et al. GC-MS analysis of cuticular lipids in recent and older scavenger insect puparia. An approach to estimate the post-mortem interval (PMI). *Anal Bioanal Chem.* 2014;406:1081–8.
 38. Moriya F, Hashimoto Y. Redistribution of methamphetamine in the early postmortem period. *J Anal Toxicol.* 2000;24:153–4.
 39. Mao S, Dong X, Fu F, Seese RR, Wang Z. Estimation of post-mortem interval using an electric impedance spectroscopy technique: a preliminary study. *Sci Justice.* 2011;51:135–8.
 40. Querido D, Pillay B. Linear rate of change in the product of erythrocyte water content and potassium concentration during the 0–120-hour postmortem period in the rat. *Forensic Sci Int.* 1988;38:101–12.