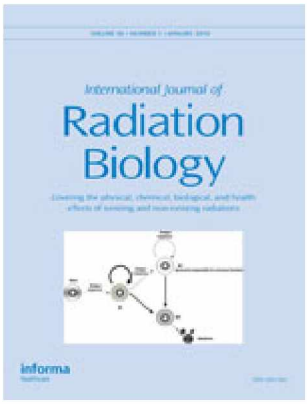
# RENEB Inter-Laboratory Comparison 2017: limits and pitfalls of ILCs.

Eric Gregoire, Joan Francesc Barquinero, Gaetan Gruel, Mohamedamine Benadjaoud, Juan Martinez Guerrero, Christina Beinke, Adyabalam Balajee, Philip Beukes, William Blakely, Inmaculada Dominguez, et al.




# RENEB InterLaboratory Comparison 2017; limits and pitfalls of ILCs.

| | |
|---|---|
| Journal: | *International Journal of Radiation Biology* |
| Manuscript ID | TRAB-2020-IJRB-0440.R2 |
| Manuscript Type: | Original Manuscript |
| Date Submitted by the Author: | n/a |
| Complete List of Authors: | GREGOIRE, ERIC; Institut de Radioprotection et de Surete Nucléaire, PSE-SANTE/SERAMED/LRACC<br>Barquinero, Joan Francesc; Universitat Autònoma de Barcelona, Biologia Animal, Biologia Vegetal i Ecologia<br>GRUEL, Gaëtan; Institut de Radioprotection et de Sûreté Nucléaire, Department of Research in Radiobioloby and Regenerative Medicine<br>Benadjaoud, Mohamedine; Institut de Radioprotection et de Surete Nucléaire, PSE-SANTE/SERAMED/LRACC<br>Martinez, Juan; Institut de Radioprotection et de Sûreté Nucléaire, Department of Research in Radiobioloby and Regenerative Medicine<br>Beinke, Christina; 4. Inst. für Radiobiologie der Bundeswehr in Verb. mit der Univ. Ulm, ; Institut für Radiobiologie<br>Balajee, Adayabalam; ORISE<br>Beukes, Philip; iThemba LABS, Radiobiology<br>Blakely, William; 43Armed Forces Radiobiology Research Institute, Uniformed Service University of the Health Sciences, Scientific Research Department<br>Dominguez, Inmaculada; University of Sevilla<br>Pham Ngoc, Duy; Nuclear Research Institute, Biotechnology<br>Monteiro Gil, Octavia; Instituto Superior Técnico, Universidade de Lisboa, Centro de Ciências e Tecnologias Nucleares<br>Güçlü, İnci; Turkish Atomic Energy Authority, Cekmece Nuclear Research and Traning Center Radiobiology Unit Yarımburgaz<br>Guogyte, Kamile; Radiation Protection Centre<br>Hadjidekova, Savina; Sofia University St Kliment Ohridski<br>Hadjidekova, Valeria; National Center for Radiobiology and Radiation Protection<br>Hande, M Prakash; National University of Singapore, Physiology<br>Jang, Seongjae; KIRAMS<br>Lumnicky, Katalin; OKK-OSSKI |

| | |
|---|---|
| | Meschini, Roberta; Universita degli Studi della Tuscia Dipartimento di Scienze Ecologiche e Biologiche, Ecological and Biological Sciences<br>Milić, Mirta; Institute for Medical Research and Occupational Health,<br>Montoro, A.; HOSPITAL UNIVERSITARIO LA FE, SERVICIO DE PROTECCIÓN RADIOLÓGICA<br>Moquet, Jayne; Public Health England Centre for Radiation Chemical and Environmental Hazards, Radiation Effects<br>Moreno, Mercedes; Hospital General Universitario Gregorio Maranon, Radiotherapy Oncology<br>Norton, Farrah; Canadian Nuclear Laboratories<br>Oestreicher, Ursula; Federal Office for Radiation Protection, Department of Radiation Protection and Health<br>Pajic, Jelena; Serbian Institute of Occupational Health, Radiation Protection Center,<br>Sabatier, L.; CEA, Radiation Biology<br>Sommer, Sylwester; ICHTJ<br>Testa, Antonella; ENEA Casaccia Research Center, Radiation Biology<br>Terzoudi, Georgia; National Centre for Scientific Research "Demokritos", Health Physics, Radiobiology & Cytogenetics<br>Valente, Marco; Institut de Recherche Biomédicale des Armées (IRBA), Département des Effets Biologiques des Rayonnements (EBR)<br>PERUMAL, VENKATACHALAM; Sri Ramachandra Institute of Higher Education and Research (Deemed to be University), Human Genetics<br>Vral, Anne; Ghent University , Basic Medical Sciences<br>Wilkins, Ruth; Health Canada, Ionizing Radiation Health Sciences Division<br>Wojcik, Andrzej; Stockholms Universitet, Molecular Biosciences;<br>Zafiropoulos, Demetre; Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Legnaro<br>Kulka, Ulrike; Federal Office for Radiation Protection, Radiation and Health |
| Keywords: | Inter Laboratory Comparison, Biodosimetry, Chromosomal aberrations, Statistical tests |
| | |

# RENEB Inter-Laboratory Comparison 2017: limits and pitfalls of ILCs.

Eric Gregoire[1*], Joan Francesc Barquinero[2*], Gaetan Gruel[1*], Mohamedamine Benadjaoud[1*], Juan S. Martinez[1], Christina Beinke[3], Adayabalam Balajee[4], Philip Beukes[5], William F. Blakely[6], Inmaculada Dominguez[7], Pham Ngoc Duy[8], Octávia Monteiro Gil[9], Inci Güçlü[10], Kamile Guogyte[11], Savina Petrova Hadjidekova[12], Valeria Hadjidekova[13], Prakash Hande[14], Seongjae Jang[15], Katalin Lumniczky[16], Roberta Meschini[17], Mirta Milic[18], Alegria Montoro[19] Jayne Moquet[20], Mercedes Moreno[21], Farrah N Norton[22], Ursula Oestreicher[23], Jelena Pajic[24], Laure Sabatier[25], Sylwester Sommer[26], Antonella Testa[27], Georgia Terzoudi[28], Marco Valente[29], Perumal Venkatachalam[30], Anne Vral[31], Ruth C. Wilkins[32], Andrzej Wojcik[33], Demetre Zafiropoulos[34], Ulrike Kulka[23+].

* These authors contributed equally to this work
\+ Chair of RENEB e.V.

1. Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses, France
2. Universitat Autonoma de Barcelona, Barcelona, Spain
3. Bundeswehr Institute of Radiobiology affiliated to the University of Ulm, Munich, Germany
4. Oak Ridge Institute for Science and Education (ORISE), USA
5. NRF iThemba LABS, Cape Town, South Africa
6. Armed Forces Radiobiology Research Institute, Uniformed Service University of the Health Sciences, Bethesda, USA
7. University of Sevilla, Sevilla, Spain
8. Center of Biotechnology, Nuclear Research Institute, Nuclear Research Institute, Dalat city, Vietnam
9. Centro de Ciências e Tecnologias Nucleares, Instituto Superior Técnico, Universidade de Lisboa, Bobadela-LRS , Portugal
10. Turkish Atomic Energy Authority, Cekmece Nuclear Research and Training Center Radiobiology Unit Yarımburgaz, Istanbul, Turkey
11. Radiation Protection Center, Vilnius, Lithuania
12. Medical University of Sofia, Sofia, Bulgaria
13. National Center for Radiobiology and Radiation Protection, Sofia, Bulgaria
14. Department of Physiology, Yong Loo Lin School of Medicine: National University of Singapore, Singapore
15. KIRAMS, Seoul, Korea
16. National Research Institute for Radiobiology & Radiohygiene, Budapest, Hungary
17. UNITUS, Viterbo, Italy
18. IMROH, Zagreb, Croatia
19. Fundación para la Investigación del Hospital Universitario LA FE de la Comunidad Valenciana, Valencia, Spain
20. Public Health England, Centre for Radiation Chemical and Environmental Hazards, Chilton, UK
21. Servicio Madrileño de Salud - Hospital General Universitario Gregorio Marañón, Madrid, Spain
22. Canadian Nuclear Laboratories, Radiobiology & Health, Chalk River, Ontario, Canada
23. Federal Office for Radiation Protection (BfS), Oberschleissheim, Germany
24. Serbian Institute of Occupational Health, Radiation Protection Center, Belgrade, Serbia
25. PROCyTOX, Commissariat à l'Energie Atomique et aux Energies Alternatives, Fontenay aux-Roses, France and Université Paris-Saclay, France
26. Institute of Nuclear Chemistry and Technology (INCT), Warsaw, Poland
27. Agenzia Nazionale per le Nuove Tecnologie, L´Energia e lo Sviluppo Economico Sostenibile, Rome, Italy
28. National Center for Scientific Research "Demokritos", NCSR"D", Athens, Greece
29. IRBA, Bretigny sur Orge, France
30. Sri Ramachandra University, Chennai, India
31. Radiobiology Research Unit, Gent University, Gent, Belgium
32. Health Canada, Ottawa, Canada
33. Stockholm University, Institute Molecular Biosciences, Stockholm, Sweden
34. Laboratori Nazionali di Legnaro – INFN, Legnaro, Italy

## Biographical notes:

**Eric Gregoire**, scientist, cytogenetician in biological dosimetry, Institute for Radiological Protection and Nuclear Safety (IRSN), Radiobiology of Accidental Exposure Laboratory (LRAcc), Fontenay aux Roses, France

**Joan-Francesc Barquinero**, PhD, Biologist, University Professor, Department of Animal Biology, Plant Biology and Ecology, Faculty of Biosciencies, Universitat Autònoma de Barcelona (UAB), Bellaterra (Cerdanyola del Vallès), Spain

**Gaetan Gruel**, PhD, Researcher and head of the Laboratory, Institute for Radiological Protection and Nuclear Safety (IRSN), Radiobiology of Accidental Exposure Laboratory (LRAcc), Fontenay aux Roses, France

**Mohamedamine Benadjaoud**, PhD, Biomathematician, Institute for Radiological Protection and Nuclear Safety (IRSN), Radiobiology of Accidental Exposure Laboratory (LRAcc), Fontenay aux Roses, France

**Juan S. Martinez**, PhD, Researcher, Institute for Radiological Protection and Nuclear Safety (IRSN), Radiobiology of Accidental Exposure Laboratory (LRAcc), Fontenay aux Roses, France

**Christina Beinke,** PhD, scientist in the cytogenetics laboratory of the Bundeswehr Institute of Radiobiology, Munich, Germany.

**Adayabalam Balajee,** Head of the Cytogenetic Biodosimetry Laboratory, Radiation Emergency Assistance Center/Training Site, Oak Ridge Institute for Science and Education, Oak Ridge Associated Universities, Oak Ridge, Tennessee, USA.

**Philip Beukes,** Radiation Protection Physicist and Head of Radiation Safety Health Environment and Quality at the National Research Foundation (NRF) iThemba LABS, Cape Town, South Africa.

**William F. Blakely,** senior scientist at his Institute and assistant professor at his University. He is a classically trained radiobiologist and for ~25 years associated with the applied biodosimetry research

and service programs. He is a member of the Scientific Research Department (SRD), Armed Forces Radiology Research Institute (AFRRI) affiliated with the Uniformed Services University of Health Sciences (USUHS), Bethesda, Maryland, United States. He is also the course director for a graduate course in Radiation Biology (PMO-582) at his University, a Council Member of the National Council on Radiation Protection and Measurements (NCRP) serving on program area committee 6 (PAC-6) entitled Radiation Measurements and Dosimetry, and members on International Standard Organization (ISO) Working Groups 18 (Performance criteria for service laboratories performing biological dosimetry by cytogenetics) and 25 (Radiological protection – Radiological monitoring for emergency workers and population following nuclear/radiological accidents – Part 1: General principles).

**Inmaculada Dominguez,** scientist, research in DNA damage and repair, lecturer in Cell Biology, Cell Culture and Radiobiological Group, Cell Biology Department, Faculty of Biology, University of Sevilla, Spain.

**Pham Ngoc Duy, PhD,** Researcher, Biodosimetry Section, Centre of Radiation Technology and Biotechnology, Dalat Nuclear Research Institute, Viet Nam.

**Octávia Monteiro Gil**, PhD, Biology- Genetic, works in the area of radiobiology and biological dosimetry, Instituto Superior Técnico, Centro de Ciências e Tecnologias Nucleares (C2TN/IST/ULisboa), Bobadela, Portugal.

**Inci Güçlü,** scientist, Head of the Radiobiology unit, Turkish Atomic Energy Authority, Cekmece Nuclear Research and Training Center, Istanbul, Turkey

**Kamile Guogyte**, PhD, Chief specialist Radiation Protection Centre Kalvarijų 153 street, Vilnius, Lithuania

**Savina Petrova Hadjidekova**, MD, Assistant Professor, Department of Medical Genetics, Medical University - Sofia, Bulgaria.

**Valeria Hadjidekova**, Director, National Center for Radiobiology and Radiation Protection, Sofia, Bulgaria

**Prakash HANDE,** Associate Professor at the Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore (NUS) and a Fellow at Tembusu College (NUS). biomarkers of radiation exposure, DNA-repair-telomeres-telomerase in ageing and cancer, experimental cancer therapeutics. Dr Hande is one of the pioneers who identified the role of DNA repair factors in telomere regulation in mammalian systems and is an expert in Radiation Biodosimetry. Dr Hande teaches cancer biology and ageing and conducts integrated study module on Biomedicine and Society and Radiation and Society. He holds adjunct professor appointments at the Vellore Institute of Technology, Vellore, India and Mangalore University, Mangalore, India. Dr. Hande is a visiting scientist at the National Institute of Radiological Sciences, Chiba, Japan. He was a consultant at the Division of Human Health, International Atomic Energy Agency, Vienna, Austria in 2015 -2016 while on sabbatical from NUS. Dr Hande is currently an expert member of the workgroup on "Biological mechanisms influencing health effects from low-dose radiation exposure" with United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR).

**Seongjae Jang**, Research scientist in biological dosimetry, National Radiation Emergency Medical Center, Korea Institute of Radiological and Medical Sciences, Seoul, Republic of Korea

**Katalin Lumniczky**, M.D., Ph.D., radiation biologist, head of the Unit of Radiation Medicine, Department of Radiobiology and Radiohygiene, National Public Health Centre, Budapest, Hungary.

**Roberta Meschini**, Research Assistant Professor, Expert in classical and Molecular Cytogenetic and Mutagenesis, Laboratory of Molecular Cytogenetic and Mutagenesis, Department of Ecological and Biological Sciences, University of Tuscia

**Mirta Milić**, scientist, molecular biologist, toxicologist and cytogenetician in biological dosimetry, Institute for Medical Research and Occupational Health (IMROH), Mutagenesis Unit, Zagreb, Croatia

**Alegría Montoro,** PhD in Biology and Master's Degree in Occupational Hazard Prevention, specialising in Hygiene and Security. Supervisor for Radioactive Facilities and head of the biodosimetry laboratory in the Radiation Protection Service (RPS), University-Polytechnic Hospital La Fe,Valencia, Spain

**Jayne Moquet,** Principal Radiation Protection Scientist in the Cytogenetics and Pathology Group, Public Health England - Centre for Chemical Radiation and Environmental Hazards, Oxfordshire, UK.

**Mercedes Moreno Domene**, MSc in radiation biology, Biological dosimetry laboratory. Radiopeathology Centre, Service of Radiation Oncology. Hospital General Universitario Gregorio Marañón (HGUGM-SERMAS), Madrid, Spain

**Farrah N Norton,** Research Scientist-Biologist, lead of the Biodosimetry emergency response capability as well the portfolio lead for the Emergency Response suite of research projects in the Safety and Security program at Canadian Nuclear Laboratories (CNL) in Chalk River, Ontario, Canada.

**Ursula Oestreicher**, PhD, biologist and head of section: "Biological Dosimetry" at the Federal Office for Radiation Protection (BfS), Oberschleissheim, Germany

**Jelena Pajic**, doctor of medical sciences, employed at the Cytogenetic Biodosimetry Laboratory, Serbian Institute of Occupational Health. Main area of research: radiation biology, biodosimetry, genotoxicology.

**Laure Sabatier**, PhD, research director, radiobiologist with molecular cytogenetics expertise, coordinator of biology and health programs and infrastructures at the fundamental research division of the French Alternative Energies and Atomic Energy Commission (CEA)

**Sylwester Sommer**, PhD, radiobiologist, Institute of Nuclear Chemistry and Technology (INCT), Unit: Radiobiology and Biological Dosimetry, Warsaw, Poland

**Antonella Testa,** Antonella Testa, radiobiologist, Italian National Agency for New Technologies, Energy and Sustainable Economic Development(ENEA), Department for Sustainability, Division Health Protection Technologies, Laboratory Health and Environment, Rome, Italy

**Georgia Terzoudi**, physicist and radiobiologist, is Director of Research at the Institute of Nuclear and Radiological Sciences & Technology, Energy & Safety, National Centre for Scientific Research

"Demokritos", working in the Health Physics, Radiobiology & Cytogenetics Laboratory in Athens, Greece.

**Marco Valente**, PhD, biologist, cytogenetician, French Armed Forces Biomedical Research Institute (IRBA), Lab: Biological Dosimetry Lab (LDBI), Brétigny-sur-Orge, France.

**Venkatachalam PERUMAL**, Professor in Human Genetics, who is having an extensive background in Radiation Genetics, with explicit training and capability in radiation biodosimetry, bystander response, and genomic instability of high and low dose ionizing radiation, differed in their LET. Sri Ramachandra Institute of Higher Education & Research, Chennai, INDIA.

**Anne Vral**, PhD, full professor and head of the radiobiology research group, principal investigator of the radiobiology group and has 30 years of experience in the field of basic and medically applied radiobiology, radiation protection, biological dosimetry and cancer. The topics related to cancer are dealing with radiosensitivity and DNA repair. An important line of research involves the development and validation of biomarkers of exposure and individual radiosensitivity. Ghent University, Belgium.

**Ruth C. Wilkins**, Research Scientist, Radiobiologist, Health Canada, Ionizing Radiation Health Sciences Division, Ottawa, Canada

**Andrzej Wojcik**, PhD, is professor of radiation biology at the Stockholm University (Sweden) and Jan Kochanowski University in Kielce (Poland). Wojcik focuses on studying cellular effects of radiation, with special focus on factors influencing the radiosensitivity and on combined exposure to radiations of different qualities.

**Demetre Zafiropoulos**, PhD in Biological Dosimetry, italian delegate, of the NEA-OECD Committee on Radiological Protection and Public Health (CRPPH), Radiation Protection Service of Laboratori Nazionali di Legnaro of National Institute of Nuclear Physics.

**Ulrike Kulka**, PhD in biology, head of section national and international cooperation and reporting at the Federal Office for Radiation Protection (BfS) in Oberschleissheim, Germany and chair of RENEB e.V.

## ABSTRACT

### Purpose

In case of a mass-casualty radiological event, there would be a need for networking to overcome surge limitations and to quickly obtain homogeneous results (reported aberration frequencies or estimated doses) among biodosimetry laboratories. These results must be consistent within such network. Inter-laboratory comparisons (ILCs) are widely accepted to achieve this homogeneity. At the European level, a great effort has been made to harmonize biological dosimetry laboratories, notably during the MULTIBIODOSE and RENEB projects. In order to continue the harmonization efforts, the RENEB consortium launched this intercomparison which is larger than the RENEB network, as it involves 38 laboratories from 21 countries. In this ILC all steps of the process were monitored, from blood shipment to dose estimation. This exercise also aimed to evaluate the statistical tools used to compare laboratory performance.

### Materials and Methods

Blood samples were irradiated at three different doses, 1.8, 0.4 and 0 Gy (samples A, C and B) with 4-MV X-rays at 0.5 Gy $min^{-1}$, and sent to the participant laboratories. Each laboratory was requested to blindly analyze 500 cells per sample and to report the observed frequency of dicentric chromosomes per metaphase and the corresponding estimated dose.

### Results

This ILC demonstrates that blood samples can be successfully distributed among laboratories worldwide to perform biological dosimetry in case of a mass casualty event.

Having achieved a substantial harmonization in multiple areas among the RENEB laboratories issues were identified with the available statistical tools, which are not capable to advantageously exploit the richness of results of a large ILCs. Even though Z- and U-tests are accepted methods for biodosimetry

ILCs, setting the number of analyzed metaphases to 500 and establishing a tests' common threshold for all studied doses is inappropriate for evaluating laboratory performance.

Another problem highlighted by this ILC is the issue of the dose-effect curve diversity. It clearly appears that, despite the initial advantage of including the scoring specificities of each laboratory, the lack of defined criteria for assessing the robustness of each laboratory's curve is a disadvantage for the "one curve per laboratory" model.

Conclusions

Based on our study, it seems relevant to develop tools better adapted to the collection and processing of results produced by the participant laboratories. We are confident that, after an initial harmonization phase reached by the RENEB laboratories, a new step towards a better optimization of the laboratory networks in biological dosimetry and associated ILC is on the way.

## 1- INTRODUCTION

In case of a large-scale radiation emergency, the dose estimation of the victims should be assessed as fast and accurate as possible. Firstly, a triage should be performed by qualified medical staff according to clinical signs. Subsequently, a categorization of the exposed people by dosimetry must be carried out according to their degree of suspected overexposure. Biological dosimetry rapid assessment complements the clinical triage by categorizing potentially exposed victims in different ranges of exposure (Vaurijoux et al. 2015; Ainsbury et al. 2014) and is a key element when physical dosimetry is not available (Christie et al. 2010; Romm et al. 2014a). Because triage based on manual dicentric chromosome analysis (DCA) is done with a low number of analyzed cells (usually 50), it makes it highly imprecise as it has large confidence intervals. For this reason, the categorization should be defined by dose ranges and not in terms of dose alone. Furthermore, after initial triage, dose assessment is needed in order to confirm the categorization and to give a more precise individual dose estimation (Romm et al. 2014a). As precise dose assessment requires the analysis of a large number of cells, usually from 500 to 1000 by DCA, the time needed for a correct dose assessment is larger than that needed for triage. As an example, using manual scoring, one operator would perform triage in 1 hour per sample, but dose estimation would take approximately two days.

In general, biodosimetry laboratories can manage only a limited number of victims at one time. Thus, in the case of a mass-casualty radiation incident, where the management of several hundreds of victims would need to be performed, there is a prerequisite for national and/or international networking. However, networking must be based on the ability to provide homogeneous results (Voisin 2015; Kulka et al. 2015, 2017). This means that for any single case, the reported chromosomal aberration frequencies or estimated doses should be consistent and comparable among the laboratories responding to the emergency. Harmonization needs standardized procedures; this is an essential point for the successful coordination of different laboratories (Beinke et al. 2013; ISO 19238; Wilkins et al. 2008; Christie et al. 2010; Beinke et al. 2011). The strategy of establishing a cooperative network among laboratories requires that each laboratory follows internationally accepted methods for analysis (IAEA 2011, ISO 19238 2014) and regular inter-laboratory comparisons (ILCs) to test performance analysis (Wilkins et

al. 2008, Di Giorgio et al. 2011). Nowadays it is widely accepted that networking should include regular international ILC exercises simulating different scenarios, as this would guarantee a more rapid response and a higher reliability of dose estimates (Wojcik et al. 2010).

During the last decade several ILCs have been performed. Some of them were focused on the triage (Wilkins et al. 2011; Lloyd et al. 2000; Ainsbury et al. 2009; Garcia et al. 2013; Romm et al. 2011, 2014a, b, Oestreicher et al 2017) while others mainly on dose-assessment (Yoshida et al. 2007; Pan et al. 2019; Bakkiam et al. 2015; Liu et al. 2016, Roy et al. 2004). In a large-scale ILC involving 7 countries from the Latin American Biological Dosimetry Network (LBDNet) and 6 laboratories from the European Union, a good agreement among participants was shown in terms of the reported dicentric chromosome yields and assessed doses. In this ILC the results after the analysis of 50, 100 or 500 cells from shared stained slides were evaluated by using robust methods described in different ISO standards (Di Giorgio 2011). Another effort in validating international networking using the DCA in the case of a potential mass-casualty event was done by Wilkins et al. (2008). Several ILCs based on triage have shown that more than 90% of the participant laboratories correctly categorize the tested samples (Miller et al. 2007; Di Giorgio et al. 2011; Beinke et al. 2011, 2013; Bhavani et al. 2014; Yoshida et al. 2007; Roy et al. 2004).

At European level, different projects have been founded in order to improve standardization and harmonization for the different biomarkers of dose. MULTIBIODOSE helped in defining what would be the best assay to use depending on different exposure scenarios (Jaworska et al. 2015, Ainsbury et al. 2014). In addition, a NATO project studied the possibility of reducing the number of analyzed cells from 50 to 20 for triage purposes (Beinke et al. 2013). Recently, several RENEB (Realising the European Network of Biodosimetry) project training sessions and ILCs have allowed the main cytogenetic assays to be homogenized and standardized among participants. Therefore, RENEB has helped in creating an efficient European network of biodosimetry laboratories (Kulka et al. 2012). The harmonization and the quality of the results for triage mode obtained among the RENEB members let us claim that at the present day RENEB is able to categorize a large number of victims in mass-casualty radiological events (Kulka et al. 2017; Gregoire et al. 2017; Oestreicher et al. 2017).

As indicated above, an individual dose assessment should be performed after triage (Romm et al. 2014a, Wojcik et al. 2010). The present ILC concerned different biomarkers (Micro-Nucleus (MN), Premature Chromosome Condensation (PCC), gamma H2AX and Gene Expression) and its main goal is to delve deeper and check the entire process needed for proper networking, from blood sample shipment to dose estimation. The present work has been focused on the DCA. In addition to the 20 RENEB laboratories, another 19 laboratories were invited to participate. Finally, as this ILC evaluates the entire process for dose-assessment by biodosimetry, we will take advantage of the large data set to critically review the statistical tools used to evaluate laboratory performance.

## 2- MATERIALS and METHODS

### a. Irradiation and Shipment

A 420 ml blood sample from a female donor (Etablissement Français du Sang (EFS), France; Agreement CPSL UNT N°13/EFS/123) was irradiated at 37°C in a water bath with 4-MV X-rays delivered by a Linear medical accelerator (Elekta Synergy, IRSN, Fontenay aux Roses, France) at 0.5 $Gy \cdot min^{-1}$, dose in water. The irradiation field was 30 x 30 cm and the distance between the source and the sample was of 1.07 m. Radiation field mapping and dosimetry was confirmed using cylindrical ionization chamber (0.125cc n° 4920) calibrated in dose to water. The blood sample was placed in 3 tubes corresponding to the different dose points, a high dose of 1.8 Gy, a low dose of 0.4 Gy, and a sham-irradiated sample. After irradiation, samples were maintained 2 h at 37°C and then the blood was aliquoted into 2 mL tubes. Blood samples were then coded as follows: the high dose as A, the low dose as C and the sham irradiated as B. Then, samples were sent to the 39 participant laboratories from 19 countries who were informed by e-mail of the shipment of three samples. The e-mail informed that there were three blind samples, that corresponded to high-, low- and sham- irradiated samples. In the same e-mail the RENEB standard scoring sheet for dicentrics, or dicentrics plus centric ring, analysis was attached.

Shipment was performed using commercial express delivery services as UN 3373 Biological Substance Category B, as described in detail in the manual of the International Atomic Energy Agency (IAEA

2011). Each package of blood samples included a temperature logger and a dosimeter to monitor the temperature and any dose received during transportation. A second blood sample (male donor, Etablissement Français du Sang (EFS), France; Agreement CPSL UNT N°13/EFS/123) was processed as above and sent to 2 laboratories for whom there were shipment issues (see section 3).

In this study, the laboratories classified as RENEB laboratories (L1 to L20) correspond to those belonging to the RENEB project that took part in the last RENEB ILC in 2015 (Oestreicher et al. 2017). The other participants are classified as non-RENEB group (L21 to L38).

**b. Cell culture and dicentric chromosome assay**

Thirty of the participant laboratories were requested to set up lymphocyte cultures. Blood samples were transmitted to three other participants by an intermediary laboratory in Bulgaria or South Korea. Thus 33 laboratories received blood samples. In all cases, cultures were processed using each laboratory's standard protocol following the recommendations of the IAEA (2011) and the ISO standard 19238 (2014). Finally, a contact laboratory from Canada set up the lymphocyte cultures and sent stained slides to its network of 6 laboratories. In all cases, the analyses were performed according to a RENEB standard scoring sheet for the dicentric chromosome assay that was provided to the 39 participants. For each sample, manual scoring of dicentric chromosomes (or dicentric chromosomes plus centric rings) in 500 cells by two different scorers if possible and using at least two slides (250 cells in each) was requested. In addition to dicentric frequency per metaphase and dose assessment (Gy) for each sample, participants were asked to report the Colcemid treatment used and the coefficients and associated errors of their calibration curve. All participants sent the results directly (30 laboratories) or indirectly (through their reference laboratory in Bulgaria, Canada and South Korea) to the organizing laboratory at the IRSN.

**c. Dose assessment**

For dose assessment, laboratories converted the frequency of aberrations observed per metaphase into absorbed dose using their own calibration curves based on dicentric chromosomes or dicentric chromosomes plus centric rings scoring. Some of the laboratories without their own calibration curve decided to use the calibration curve data available in the last technical IAEA report on biodosimetry

(IAEA, 2011). The calculations needed to convert the observed dicentric chromosomes (or dicentric chromosomes plus centric rings) into estimated doses were made by means of various software programs: CABAS V2.0 (Deperas et al 2007), different versions of Dose Estimate (Ainsbury and Lloyd 2010), or in-house developed software based on Microsoft Excel (L6, L15, L19 and L34). All laboratories estimated the doses in Gray (Gy) and included the 95% confidence intervals as requested.

**d. Statistics**

To assess the performance of each laboratory and the reproducibility of the exercise, the statistical analysis followed the (ISO 5725 1998) recommendations which provide detailed guidance of general statistical methods to use in proficiency testing schemes. These methods were successfully applied for biological dosimetry in Di Giorgio et al. (2011). In brief, the robust estimations of the mean and standard deviations of frequencies or doses were performed using the Algorithm A (algA function of the R software "metrology" package) (ISO 13528, 2015). This algorithm yields robust location and scale estimates by the "winsorisation" of the original data (the extreme values, instead of being deleted, are shifted towards the bulk of the data using adequate upper and lower thresholds obtained by an iterated scale). The "Breakdown points" for these estimators (proportion of outliers without an adverse impact on the estimates) are approximately 30, which constitutes an adequate resistance to outlying values. A robust estimation of the coefficient of variation can then be obtained as the ratio of the robust standard deviation to the robust mean.

Once the mean and standard deviation robustly estimated, the performance analysis was conducted using the Z- and U-tests. The Z-test measures the deviation of each laboratory's reported frequency or estimated dose from the robust mean of the reported frequencies or the delivered dose, both considered as reference values. The Z-test also takes into account a robust standard deviation from the reported frequencies or doses, and a standard uncertainty of the reference value. Laboratory performance using the Z-test categorizes reported values into "satisfactory" when the $|Z|$ value is $\leq 2$, "questionable" for a $|Z|$ value between 2 and 3, and "unsatisfactory" when the $|Z|$ value is $\geq 3$. Z-tests do not consider the uncertainty of each participating laboratory. On the other hand, the U-test considers the mean value and its confidence interval. With the U-test, the results of each laboratory are interpreted considering the

upper critical value of Student's t distribution, usually with a 0.05 probability of exceeding the critical value, and with N-1 degrees of freedom (where *N* is the number of laboratories). For both tests, Z and U, and to prevent against the multiple testing issues in the statistical inference, the Benjamini-Hochberg (BH) (Benjamini and Hochberg 1995) adjustment was performed for controlling the false discovery rate (FDR). This FDR-based control has been widely used in cases where a large number of hypotheses are simultaneously tested and has been shown to be less conservative than the Bonferroni adjustment (Shaffer 1995).

## 3- RESULTS

### a. Shipment

A total of 39 laboratories were involved in this ILC. From the initial shipping of blood samples by the organizing laboratory (IRSN), 11 laboratories belonging to the European Union (EU) received them within a period of 24h; and 22 laboratories received them within a period of 48 h regardless of location (EU or non-EU). 4 laboratories received the blood samples after a period larger than 48 h (from 48.5 h to 68 h), and 2 laboratories did not receive the samples due to an issue with customs authorities. A new shipment for the latter 2 was made by the IRSN laboratory and it was received within 48 hours. Each package included a dosimeter and none of the recorded doses was above 0.1 mSv. Among the laboratories that received blood samples, 32 out of 33 were able to set up lymphocyte cultures and to successfully obtain chromosome spreads. Considering all participants, including the 6 labs that received coded slides with chromosome spreads, a total of 38 laboratories were able to report their results.

### b. Reported Frequencies

Table 1 shows the total number of dicentric chromosomes or dicentric chromosomes plus centric rings found by each laboratory for the three evaluated samples (A, B and C). From the total 114 reported values (38 labs x 3 samples), 99 of them correspond to the analysis of around 500 cells and the other 15 values correspond to 300 analyzed cells or less (Table 1). Two laboratories submitted 2 dose-effect curves each for manual scoring, as follows: one sent curves based on different kinds of staining (Giemsa (L2) or FISH coupled with pan-telomeric and pan-centromeric probes (L2b)), and one sent curves based on chromosomal aberration scoring (dicentrics (L31) and dicentrics plus centric rings (L31b)). All the sent results have been included in the analysis to show what can happen in a real case. Indeed it is possible, whatever the cause, that a laboratory obtains a poor mitotic index, even when it is a very experienced laboratory.

TABLE 1NEAR HERE

The figure 1 shows only the dicentric frequency per metaphase in order to effectively compare the same frequencies to each other. For sample A (1.8 Gy), the dicentric frequencies sent by each participant laboratories are shown in Figure 1A. The observed frequencies of dicentric chromosomes per cell ranged from 0.10 to 0.34. The robust estimate of the mean (± robust standard deviation) was 0.22 ± 0.058, and the coefficient of variation (CV) was of 27%. When only RENEB laboratories were considered CV was 20%. Figure 1 also shows the results of the Z and U tests.

FIGURE 1 NEAR HERE

Z- and U-tests were only done using the frequency of reported dicentric chromosomes per metaphase (Figure 1). For both tests, the BH adjustment was applied. The Z-score obtained for 97% of the labs were satisfactory. Different results were obtained using the U test, where the results of 10 labs (26% of labs) were unsatisfactory after BH adjustment. Evaluating separately RENEB (from L1 to L20) and non-RENEB laboratories (L21 to L38), we can notice that 85% of RENEB laboratories had satisfactory U-test values and among non-RENEB laboratories, only 61% showed satisfactory U-scores.

For sample B (0 Gy), dicentric chromosome frequencies sent by each participant laboratories are shown in Figure 1B. The observed frequencies of dicentric chromosomes per cell ranged from 0.0 to 0.01 and the robust estimate of the mean (± robust standard deviation) was 0.0014 ± 0.0017 and the CV was 128%. The results of the Z and U tests are also shown in Figure 1B. Concerning the Z-test, the results of only three participants (L14, L20 and L29) were considered unsatisfactory. 90% of the RENEB and 94% of the non-RENEB participants had a satisfactory Z-score. The U- test considered that all the values given by the laboratories are satisfactory.

The dicentric frequencies for sample C (0.4 Gy) sent by each participant laboratories are shown in Figure 1C. Observed mean frequencies of dicentric chromosome per cell ranged from 0.0 to 0.08 and the robust estimate of the mean (± robust standard deviation) was 0.025 ± 0.011. The CV was 44%.

As shown in Figure 1C, only one Z value was considered as unsatisfactory (L7). Therefore, 97% of the participants present satisfactory Z-scores. When the U-test was applied, 2 frequencies gave unsatisfactory results (L36 and L37). With the U-test, L7 was not unsatisfactory anymore as the uncertainty associated with its frequency per metaphase is quite large due to the analysis of only 12 cells. So, 95% of the participants had a satisfactory U-score.

### c. Estimated Doses

TABLE 2 NEAR HERE

The second step of this intercomparison was to estimate the three delivered doses. The ILC requested each laboratory to calculate the estimated doses and their associated confidence intervals, using their own dose-effect curve and applying the statistical method established in their laboratory. In addition, the RENEB scoring sheet requested each laboratory to indicate the coefficients and standard errors of the calibration curve used (Table 2). Twenty-nine laboratories sent the coefficients of a single dose-effect curve, generally constructed using gamma- or X-rays (Table 3). Four laboratories submitted 2 dose-effect curves each, as follows: 2 participants sent curves based on different irradiation sources (gamma- (L4 and L5) and x-rays (L4b and L5b)), and the two others (L2/L2b and L31/L31b) were mentioned earlier in the "reported frequencies" section. Finally, five laboratories did not have any calibration curve but two of them chose to use the calibration curve data available in the last technical IAEA report on biodosimetry (IAEA, 2011), as it is also an acceptable method. For the other three, as it was their first time of participation to an intercomparison, they were not aware of the possibility to use an established dose-effect curve. Figure 2 shows only the values from the participants that reported an estimated dose.

TABLE 3 NEAR HERE

FIGURE 2 NEAR HERE

Using the values sent by each laboratory, 65% of the reported dose estimation participants include the high dose (sample A) in the 95% confidence interval of their dose estimates (93% and 89% for sham

irradiation (sample B) and low dose (sample C), respectively). As mentioned above, each laboratory calculated the absorbed doses using the program routinely used in their laboratory, and there was a great heterogeneity in the calculation of the 95% confidence interval. In fact, 16 laboratories used the CABAS software that only considers Poisson's error on the observed yield. 13 participants used Dose Estimate software, that can consider both, the error of the curve and the error of the observed yield of dicentric chromosomes applying the delta method (IAEA manual 405, 2001). Among the 13 participants which used the Dose Estimate software, 11 considered the delta method, and 2 only considered the error of the observed yield. As well 6 laboratories that used their own software applied a Merkle approach to consider both errors (Merkle, 1983). Finally, 3 laboratories gave no results on dose estimation. Among the 38 laboratories that sent results, 2 laboratories sent miscalculated doses due to typo errors. To avoid the impact of this heterogeneity in Z- and U-test analysis, all the dose estimations were recalculated using a single method. The method used was Merkle's approach that was proposed in the last IAEA manual (IAEA, 2011). However, because covariances of the fitted coefficients of curves were not previously requested, the 95% confidence intervals were calculated considering only the standard errors on curve coefficients. These results are reported in Figures 3, which show 39 results each because some laboratories provided 2 dose effect curves leading each to dose estimations.

FIGURE 3 NEAR HERE

For the high dose (sample A), the estimated doses ranged from 1.31 to 2.51 Gy, and 90% of the participants included the delivered dose in their 95% confidence interval (Figure 3A). After applying the BH adjustment all laboratories showed satisfactory Z-scores. Using the U-test, 90% of the laboratories showed satisfactory results, 96% for RENEB and 81% for NON-RENEB participants. The CV was of 15%.

For the sham-irradiated sample (B), the estimated doses ranged from 0.0 to 0.19 Gy, and in all cases the 95% confidence intervals included the 0 Gy dose, except L35 (Figure 3B). For sample B, results cannot be analyzed using the Z-test because of the algorithm A convergence failure of the robust

standard deviation estimation (abundance of null values). The U-test showed that 97% of the results were satisfactory. All RENEB and 94% of non-RENEB participants had satisfactory U-scores.

Results for the low dose (sample C) can be seen in Figure 3C. Estimated doses ranged from 0.24 to 1.20 Gy, and the Z-test shows unsatisfactory result for only one participant. Therefore, 97% of the laboratories had satisfactory scores. All U-scores were satisfactory. The CV was 29%.

Importantly, we noticed a substantial heterogeneity in the calibration curves from the participants as reported in Table 3 and Figure 4. Concerning the gamma-rays calibration curves, the lowest dose rate was 0.04 $Gy \cdot min^{-1}$ (L3), and the highest 1.16 $Gy \cdot min^{-1}$ (L34). For calibration curves using X-rays, dose rates ranged between 0.35 $Gy \cdot min^{-1}$ irradiating with X-rays of 243 kVp (L4b) and 2.5 $Gy \cdot min^{-1}$ irradiating with X-rays of 6 MeV (L19). The IAEA technical report (IAEA, 2011) recommends that to produce a dose-effect curve applicable to an acute accidental exposure the dose rate should be chosen such that all doses are given in less than 15 min. Considering this recommendation and taking into account that usually the highest dose used in a calibration curve is 4 or 5 Gy, a dose rate of about 0.34 $Gy \cdot min^{-1}$ will allow to irradiate the highest dose in less than 15 minutes. Therefore 7 laboratories (L3, L7, L16, L22, L33 and L34) used a dose rate that is under the IAEA recommendations.

An alternative to reporting satisfactory result rates is to rank the results of each laboratory belonging to the same network for a given sample based on their Z- and U-scores. Table 4 shows the differences in ranking of the laboratories between the 2 tests.

TABLE 4 NEAR HERE

In fact, differences in scoring criteria should be balanced by the use of individual curves, which logically includes the specific scoring criteria of each laboratory. This effect is not clearly observed in the present study. Table 5 shows the differences of laboratory ranking from the frequency to the dose estimation by Z-score. For example, at group level, the mean Z-score calculated for the RENEB network or the non-RENEB group does not change as much between frequency and estimated dose. For frequency and dose estimation, RENEB and non-RENEB laboratories were ranked based on their Z-scores, from the lower

to the higher values. The mean of the rank obtained for RENEB and non-RENEB laboratories are 17.4 and 21.8 respectively. This is quite the same for dose estimation, the mean of the laboratory rank based on Z-scores is 18.7 for RENEB, and 21.8 for non-RENEB laboratories. It should be noted that the mean dose for all laboratories is not far from the delivered dose (1.74 Gy vs 1.80 Gy).

TABLE 5 NEAR HERE

The curves reported for the present ILC show great variability in their calibration curve coefficients (Table 2) and highlight the existing diversity among laboratories. A more visual representation of these differences can be seen in figure 4. As an example, a frequency of 0.5 dicentric chromosome per metaphase gives a dose of 1.80 Gy for L2b and a dose of 3.90 Gy for L9.

FIGURE 4 NEAR HERE.

## 4- DISCUSSION

In a mass-casualty radiological event, networks of biological dosimetry laboratories can decide to share different types of samples such as whole-blood, fixed cells, slides or metaphase images. Multiple Inter-Laboratory Comparisons (ILCs) have already tested different possibilities: blood (Roy et al. 2004; Oestreicher et al. 2017, Bakkiam et al. 2015, Romm et al. 2011, Pan 2019, and Wilkins et al. 2008); fixed cells (Roy et al. 2004); slides (Liu et al. 2016; Miller et al. 2007); or metaphase images (Garcia et al. 2013, Romm et al. 2014a and 2016). The present ILC has chosen to send whole blood to test all the steps of a biological dosimetry study (i.e. blood culture, slide preparation and staining, dicentric analysis and dose estimation).

Evaluating the shipment, 87% of the participant laboratories received the blood samples within 48h, including those outside Europe (Canada, USA, South Africa, South Korea, India and Vietnam). In addition, 97% of the laboratories were able to obtain chromosome spreads, even those that received the samples after 48h. In fact, only 6 out of 38 laboratories did not reach the 500 metaphases needed. The only laboratory that did not obtain any chromosome spreads received the blood sample in 48h. Therefore, no link could be established between sample-travel time and culture growth, and some delay in the shipment did not prevent lymphocyte growth in this study. In future ILCs it would be of interest to report the mitotic index in order to evaluate lymphocyte activation. Moreover, the impact of the shipment itself has been tested in other exercises. Particularly, in the ShipEx exercise between the Latin-American network (LDBNet) and several laboratories around the world. In this case, blood samples were also properly received and lymphocytes were able to satisfactorily grow for most of the participants (Garcia et al. 2013). The blood shipment has also been tested in other European ILCs (Beinke et al. 2013), where the same observation was made for the longest shipment times, including 96 h but this was not optimal (Oestreicher et al. 2017).

In our study, the dose received during the transport of the samples (mainly cosmic radiation and X-ray safety checks) amounted to a maximum of 0.1 mSv. In addition, only long-distance shipments were exposed to measured doses between 0.05 and 0.1 mSv. As comparison, similar exposures were reported in the ShipEx-1 exercise (Garcia et al. 2013). Thus, these dose levels can be considered negligible

compared to the sensitivity limits of the biological dosimetry method used. However, it seems prudent to systematically include a dosimeter in the blood sample shipment, in case there are abnormal exposure levels during transit safety checks. The results presented here and those previously reported (Oestreicher et al. 2017, Wilkins et al. 2008) show that blood samples could be shared among laboratories around the world in the event of a major radiological accident in order to perform biological dosimetry based on chromosome aberrations.

**Interpretation of ILC Results**

Periodic ILCs allow the evaluation of the performance of laboratories that belong to a network. They help to standardize practices and contribute to the improvement of the quality and robustness of the results from such a network. One important aim of ILCs is to identify problems encountered by the participants and define actions for improvement, such as harmonization, training and dose estimation exercises. In biological dosimetry, the results are mainly based on the estimation of the chromosome aberration frequency per metaphase. This value is subsequently converted into an estimated absorbed dose using a pre-established dose-response curve specific to each laboratory. For biodosimetry laboratories, the main goal of an ILC is to compare the results for these two values, frequency and estimated dose, among the participant laboratories.

The objectivity of these comparisons is generally achieved through the Z- and U- score tests (Di Giorgio et al. 2011). These two quantities evaluate, under different normalizations, the difference between the value reported by each laboratory and a reference value considered as correct (i.e. the robust mean for frequencies, or the delivered dose for dose estimation). In fact, the Z-Score is computed under a common normalization based on the robust standard deviation while the U-score is computed using a laboratory specific normalization based on the uncertainty measurement of each participant (which is highly associated to the number of cells scored, but also to the level of exposure). Thus, these two tests give complementary elements to interpret ILC results. In the present study, the U- and Z-scores were adjusted using the Benjamini & Hochberg (1995) correction in order to take into account the large number of calculated scores (at least one for each participating laboratory) and the associated increase of false

positive risk. Finally, by defining thresholds, one could distinguish acceptable, questionable and unsatisfactory results. The comparison of frequencies aims to provide an overview of the state of harmonization between the participating laboratories concerning chromosome aberration recognition. In other words, this analysis allows the evaluation of the homogeneity among participating laboratories concerning aberration detection and scoring criteria. From an overall perspective, with the Z-score, the percentage of satisfactory results decreases with the level of exposure: 100%, 97% and 92% for samples A (1.8 Gy), C (0.4 Gy), and B (non-exposed) respectively. Contrary to what the percentages of satisfactory results might suggest, it cannot be concluded solely on the basis of the Z-score that laboratory harmonization is worse at low doses than at high doses. The reason is that these percentages are simply not directly comparable. In fact, the standard ISO 13528 (2015) justifies the use of the limits "2" and "3" for the Z-score by the fact that "measurements that are carried out correctly are assumed to generate results that can be described by a normal distribution". Therefore, it is easy to see that the validity of the Z-scores limits (2 and 3) is intrinsically related to the large-sample asymptotic normal approximation of a Poisson distribution, which is usually used to describe the distribution of dicentric chromosomes in a uniform irradiation context. The Berry-Essen Theorem (Berry 1941, Essen 1942) provides an easy way to quantify this convergence rate which, in the case of a Poisson distribution, states that a bound on the maximal error of the normal approximation is inversely proportional to the square root of the product of the number of metaphases times the dicentric rate.

According to the sample A and the sample C aberration rates per cell (approximately 0.2 and 0.02 respectively), this implies that 5000 analyzed metaphases are needed for the low dose (sample C) to achieve the normal approximation precision after analyzing 500 metaphases of the high dose (sample A). In other words, by fixing the number of analyzed metaphases to 500 for all investigated doses, the corresponding Z-score distributions are significantly different in terms of normal approximation, making it inappropriate to have common satisfactory/unsatisfactory thresholds (here 2 and 3).

The same conclusion can be made for the U-score, even though it gives opposite results. With the U-test, the percentage of satisfactory results decreases when the sample dose increases: 100% for non-irradiated sample (dose B), 95% for 0.4 Gy (dose C) and 74% for 1.8 Gy (dose A). Once again, and for

the same reasons explained above for the Z-scores, it would be erroneous to conclude that laboratory harmonization is worsening as the dose to be estimated increases.

An alternative to reporting satisfactory result rates is to rank the results of each laboratory belonging to the same network for a given sample based on their Z- and U-scores. As explained above, the methodology underlying the Z- and the U-score is not the same and the analysis of the ranking obtained with each one should be interpreted in light of these differences. The Z-score ranks the laboratories based on the distance between the value reported by each of them and the reference frequency (i.e. the robust mean of all reported frequencies). Basically, the farther you are from this average value representative of the group, the lower you are ranked. A disadvantage of this method is that two laboratories that report the same mean frequency will have the same score, even if one of them has a larger uncertainty for the measurement. This can be illustrated by comparing the Z-score of L6 and L7 for sample A. The two laboratories obtained a Z score very similar (0.34 and 0.28) as their reported frequencies for sample A are similar (Table 4). However, the frequency of L7 has a much higher uncertainty due to the low number of scored metaphases, and it can be considered less reliable than the result of L6, which is not reflected in the Z-score ranking. The U-score makes it possible to account for this difference between the 2 laboratories, but not in the direction that one would expect. In fact, the U score for L7 (0.12), is lower than for L6 (0.86). Therefore, when performing the U-test for two laboratories with similar frequencies, one of them can be better ranked because of its large uncertainty. Because the number of dicentric chromosomes that can be detected will depend on the delivered dose and on the number of cells analyzed, these two tests should be used carefully when ILC frequencies of detected aberrations are considered. It seems more reasonable to use these tests to evaluate the level of harmonization between laboratories, or networks of laboratories, rather than to evaluate each laboratory's performance. The present RENEB ILC involves laboratories belonging to different groups (RENEB network and non-RENEB participants) that have independently harmonized dicentric chromosome scoring. In the present ILC, RENEB laboratories constitute half of the participants and most of them have already participated to several ILCs (Oestreicher et al. 2017, Jaworska et al. 2015, Romm et al. 2014a, Ainsbury et al. 2014). This has a strong effect in the robust mean and robust standard deviation considered as reference values.

The global analysis of frequencies by Z-score and U-score as presented in Table 4 can lead to misinterpretations. As mentioned above, the real interest of the frequency analysis is to evaluate the level of harmonization in the recognition of dicentric chromosomes. The present study involving 38 laboratories around the world that do not belong to the same network, or do not even belong to a network, necessarily highlights an expected heterogeneity in the results, which is not very surprising or informative. It might be of interest to focus on a sub-group of laboratories that have worked to harmonize themselves, and to assess the gain associated with this harmonization process.

TABLE 6 NEAR HERE

Table 6 presents the results for sample A and for the Z-score analysis performed only on the 20 laboratories belonging to the RENEB network. Within this group, the robust coefficient of variation is 20.1%, with 3 labs (15% of all RENEB labs) showing questionable results (L11, L18 and L3). If these 3 laboratories are excluded, the coefficient of variation calculated from the frequencies obtained by the remaining 85% of the laboratories is around 15%. These values can then be compared to the expected value for the coefficient of variation which can be obtained by simulating 20 or 17 chromosome aberration frequency estimates following a Poisson distribution with a parameter (lambda) equal to the robust means observed on the RENEB subgroup, and taking into account the respective numbers of metaphases scored by each laboratory. Then, the median value of these “theoretical” coefficients of variation is 13.5% with 95% confidence interval of [7.7% - 27.4%]. This means that, due to the stochastic nature of the measures, 20 laboratories involved in a "fully harmonized" ILC situation is expected to obtain, in median, a coefficient of variation of 13.5%. Thus, the dispersion of the values obtained for the RENEB network, 20% or 15% is included within the 95% confidence interval of the “theoretical” coefficient of variation and close to the median “theoretical” value of 13.5%. In comparison, the robust coefficient of variation obtained for all 38 laboratories is 26.7%, and if only the non-RENEB laboratories are considered, the dispersion reaches a value of 36.5%. This shows that an intercomparison analysis based on chromosomal aberration frequencies only makes sense among laboratories that are involved in a common effort of harmonization. This should not be interpreted as a

better proficiency of a specific network in detecting dicentric chromosomes with respect to another, but as a reflection of different ways of harmonization.

In conclusion, although Z- and U- tests are accepted methods to assess laboratory performance in metrology (ISO 13528 2015), they are not ideal for ILCs. To this day, no commonly used tool proves to be fully adapted and relevant to the needs of ILCs that are based on the frequency of radiation-induced dicentric chromosomes per metaphase. To mitigate this deficiency, it seems appropriate for the reliability of future ILCs to focus on radiation doses that are able to generate enough dicentric chromosomes for 500 analyzed cells. This would limit the impact of Poisson uncertainties on the ILC results. In addition, it seems essential to only include in the intercomparison analysis those laboratories that have analyzed the requested number of metaphases, and to exclude those that have not, thus allowing a comparison with an equivalent Poisson uncertainty. Otherwise, a comparison of results from all participants appears hazardous. Additionally, one should consider that ILCs may include laboratories from different networks that could have their own harmonized way of scoring dicentric chromosomes. This could lead to questionable or unsatisfactory results because of different scoring criteria, and it should not be interpreted as bad performance, but as a lack of harmonization among all the participating laboratories.

While ILCs based on the frequency evaluate the level of harmonization of scoring criteria, intercomparisons based on estimated doses involve additional elements to be taken into account when analyzing the results. One is the dose-effect curve required to estimate a dose from the observed frequency, of which most of the participant laboratories have their own. It is widely accepted that in ILCs, better results are generally obtained with estimated doses than with observed frequencies (Di Giorgio et al. 2011). In fact, differences in scoring criteria should be balanced by the use of individual curves, which logically includes the specific scoring criteria of each laboratory. This effect is not clearly observed in the present study. For example, at group level, the mean Z-score calculated for the RENEB network or the non-RENEB group does not change as much between frequency and estimated dose. For frequency, the mean rank based on the Z-scores obtained for RENEB and non-RENEB laboratories are 17.4 and 21.8 respectively. This is quite the same for dose estimation, the mean rank based on the Z-

scores are 18.7 for RENEB, and 21.8 for non-RENEB. This is confirmed at the laboratory level, as drastic changes in Z-score ranking (gain or loss of more than 15 places) between frequency and dose estimation are limited to a minority of laboratories (Table 5). This is the case for L18, which systematically improves its Z-score by more than 1.5 points between frequency and estimated dose (a progress of 25 ranking places). Inversely, L9 gained more than 2 Z-score points when its dose-effect curve was used to convert its chromosome aberration frequency to an estimated dose (a 35-row drop in the overall ranking).

It is interesting to note that, although these changes in results between frequency and dose are small in magnitude for most laboratories, globally, they are quite unfavorable. In fact, an increase of the Z-score is observed for the majority of participants (58% for the high dose, sample A and 55% for the low dose, sample C) when estimated doses are considered. Indeed, for the high dose, a mean loss of 2 ranks per lab were observed between the ranking obtained for frequency and the one obtained for dose estimation. This could indicate that the dose-effect curves include biases that prevent them from positively compensating for differences in scoring criteria. The curves reported for the present ILC show great variability in their calibration curve coefficients (table 2) and highlight the existing diversity among laboratories. A more visual representation of these differences can be seen in figure 4.

The above-mentioned differences have multiple origins such as the number of dose points used to calibrate the curve, the number of metaphases analyzed at each dose point, the dose-rate and the radiation source (X- or gamma-rays). Another source of uncertainty is the way that the delivered doses were calculated (Trompier et al, 2017). Briefly, depending on the radiation source, X-or gamma-rays, and their energy, the calculation of the delivered dose to the samples can be based on air Kerma or dose to water. Depending on the overall energy of the source, this could lead to different absorbed dose values for the same irradiation. Consequently, this can impact the result of the dose estimation in an ILC if the doses of a given dose-effect curve are not calibrated the same way than the dose delivered to the analyzed sample. It is important to mention that usually in biodosimetry laboratories, all these details are not very well traced. Furthermore, there are no minimum criteria for defining whether or not a dose-response curve is acceptable for use in a given intercomparison. Currently, and in most of ILCs, calibration curves from all participants are used, regardless of the way they are built. In fact, 7 laboratories reported dose

effect curves that were built using a dose rate too low to fully respect the IAEA recommendations concerning how to build a dose effect curve applicable to an acute exposure. This point must absolutely be considered for future intercomparisons as it has a very strong impact on the interpretability of the results and on the identification of improvement areas for a specific network. Additionally, evolution in the scoring criteria within a laboratory over the time elapsed between the calibration curve establishment and the present intercomparison may lead to additional uncertainty in the dose assessment. In fact, scorers are changing over the time so there is a need for periodical harmonization. Another important issue is how the participation in intercomparisons have modified the scoring criteria and so the dose-effect curve. This is particularly important if the dose-effect curves were produced prior to the harmonization work carried out within the RENEB network.

One more issue brought forward by this intercomparison was the lack of homogeneity in the calculation of the uncertainties associated to the doses reported by the participants. In fact, there are different generally accepted ways to estimate a dose and its associated uncertainties, as several calculation software programs are available (CABAS, Dose Estimate and Microsoft Excel-based spreadsheets). Considering that these tools do not implement the same methodologies to calculate uncertainties, it made it difficult to compare the raw reported values because they were not calculated in a homogenous manner. In the present study, the estimated doses and uncertainties initially sent by participants were calculated by each laboratory using their own methods. This led to a great heterogeneity in the reported values and in the reported curve coefficients, which further complicated their interpretation in the context of an intercomparison. For this reason, all dose estimates were re-calculated using the reported frequencies and their own dose-response curves using the method described by Merkle et al (1983) and mentioned in the IAEA manual (2011). For future ILCs, it seems essential to clearly define the methodology to be applied by the laboratories for the calculation of the dose and the associated uncertainties. To go further, the implementation of a single integrated and open-ended tool available to the participants seems to be relevant. This was the strategy adopted by the RENEB association, through the development of BiodoseTools, a software based on R with a Shiny interface (https://github.com/biodosetools-team/biodosetools).

## 5- CONCLUSION AND PERSPECTIVES

Standardization of chromosomal aberration scoring during the various European projects (MULTIBIODOSE and RENEB) has improved results of dose assessment in ILC exercises (Jaworska et al. 2015, Oestreicher et al. 2017). The present work demonstrates that harmonized and trained networks such as RENEB (in terms of chromosomal aberration scoring) obtain better results than a non-harmonized group. This is illustrated by the high level of satisfactory results obtained either in frequency or dose by L1-L20 when using classical intercomparison analysis tools, such as the Z-score and its associated decision thresholds. However, one cannot conclude that RENEB laboratories are fully harmonized, not only for those non-satisfactory results but also by the statistical tools used. These statistical tools appear to be limited and are not able to advantageously exploit the richness of results from large intercomparisons. At present, these tools do not allow a fine diagnosis of laboratory performance, neither do they serve as new avenues for improvement for the network of laboratories. For example, it would seem interesting to be able to easily discriminate results such as those obtained by the L5, L7 and L9, which intuitively do not seem equivalent, but are considered as such by looking at their Z- and U-scores. After this first stage of harmonization using these tools, the use of other approaches to test laboratory performance in future intercomparisons seems to be necessary. Solutions based on the bias-variance trade-off are currently being explored.

Another issue highlighted by this ILC is the question of the infinite diversity of dose-effect curves. It clearly appears that, despite the initial advantage of including the scoring specificities of each laboratory, the lack of recommendations and minimum criteria to evaluate the robustness of each laboratory's curve seems to be a negative point for the model of "a curve per lab". The construction of a robust curve is a long-term procedure, which should be part of a constant and dynamic evolution process in order to take into account the changes occurring over time in the laboratories, or the evolutions inherent to the process of harmonization of a network. In addition, the relevance of a dose-effect curve established 25 or 30 years ago by members who are no longer present in a given laboratory is questionable. By definition, the process of harmonization would generate a change in practices and may raise questions about the validity of a pre-existing dose-response curve. One of the main advantages of a large laboratory network

is its power in terms of data production. The present intercomparison generated the analysis of a striking 20,000 different metaphases per dose. When harmonization of practices is achieved, such a network could build an extremely robust dose-response curve in just 2 or 3 intercomparisons. This would also have the advantage of consolidating practices in terms of calculating coefficients and the associated uncertainties, making it a strategy that should be seriously considered in large.

Finally, and in the same spirit of unification, it seems relevant to develop tools that are better adapted to the collection and processing of results produced by the various participant laboratories. For the moment, this collection happens at a relatively small-scale (notably through the exchange of spreadsheet files). The coupling of tools such as BiodoseTools and web portals for collecting results seems to be particularly promising, both in terms of definition and application of the methodologies necessary for their processing (in particular, the calculation of uncertainties), but also in terms of the reliability associated with the traceability of results.

After a first harmonization phase lasting more than ten years (Kulka et al. 2017, Oestreicher et al. 2017, Gregoire et al. 2017), and even if there is still room for improvement, the level of harmonization reached by RENEB members definitely confirms the operational value of international networks of biological dosimetry laboratories, particularly in the case of large-scale radiological accidents.

ACKNOWLEDGEMENTS & DISCLAIMERS

One author's (WFB) efforts in this study was funded by AFRRI's intramural protocols RBB44313 and AFR-B4-4313. The author (WFB) wishes to thank Uma Subramanian and Dr David L. Bolduc for their contributions in this study. The opinions and assertions expressed herein are those of the author (WFB) and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences or the United States Department of Defense. The research protocol was reviewed and approved by the USUHS IRB Committee in accordance with all Federal regulations governing the protection of humans in research.

DISCLOSURE STATEMENT

No conflict of interest was reported by the authors

BIBLIOGRAPHY


1. Ainsbury E, Livingston GK, Abbott MG, Moquet JE, Hone PA, Jenkins MS, Christensen DM, Lloyd DC and Rothkamm K. Interlaboratory variation in scoring dicentric chromosomes in a case of partial-body x-ray exposure: implications for biodosimetry networking and cytogenetic ''triage mode'' scoring. 2009. Radiat Res. 172:746–752.
2. Ainsbury EA and Lloyd DC. 2010. Dose estimation software for radiation biodosimetry. Health Phys. 98:290–295.
3. Ainsbury EA, Al-Hafidh J, Bajinskis A, Barnard S, Barquinero JF, Beinke C, de Gelder V, Gregoire E, Jaworska A, Lindholm C, Lloyd D, Moquet J, Nylund R, Oestreicher U, Roch-Lefévre S, Rothkamm K, Romm H, Scherthan H, Sommer S, Thierens H, Vandevoorde C, Vral A and Wojcik A. 2014 Feb. Inter- and intra-laboratory comparison of a multibiodosimetric approach to triage in a simulated, large scale radiation emergency. Int J Rad Biol. 90:193–202.
4. Bakkiam D, Bhavani M, Anantha Kumar AA, Sonwani S, Venkatachalam P, Sivasubramanian K and Venkatraman B. 2015. Dicentric assay: inter-laboratory comparison in Indian laboratories for routine and triage applications. Appl Radiat Isot. 99:77–85.
5. Bhavani M, Tamizh Selvan G, Kaur H, Adhikari JS, Vijayalakshmi J, Venkatachalam P, Chaudhury NK. 2014. Dicentric chromosome aberration analysis using giemsa and centromere specific fluorescence in-situ hybridization for biological dosimetry: An inter- and intra-laboratory comparison in Indian laboratories. Appl Radiat Isot. 92:85–90.
6. Beinke C, Oestreicher U, Riecke A, Kulka U, Meineke V, Romm H. Inter-laboratory comparison to validate the dicentric assay as a cytogenetic triage tool for medical management of radiation accidents. 2011. Radiat Meas. 46:929-935.
7. Beinke C, Barnard S, Boulay-Greene H, De Amicis A, De Sanctis S, Herodin F, Jones A, Kulka U, Lista F, Lloyd D, Martigne P, Moquet J, Oestreicher U, Romm H, Rothkamm K, Valente M, Meineke V, Braselmann H and Abend M. 2013. NATO dosimetry study Laboratory Intercomparison of the Dicentric Chromosome Analysis Assay. Radiat Res. 180:129–137.

8. Benjamini Y and Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat (Methodol). 57(1):289-300.
9. Berry AC. 1941. The accuracy of the Gaussian approximation to the sum of independent variates. Trans. Amer. Math. Soc. 49(1):122-136.
10. Christie DH, Chu MC and Carr Z. 2010. Global networking for biodosimetry laboratory capacity surge in radiation emergencies. Health Phys. 98:168–17.
11. Deperas J, Szluinska M, Deperas-Kaminska M, Edwards A, Lloyd D, Lindholm C, Romm H, Roy L, Moss R, Morand J, Wojcik A. 2007. CABAS: a freely available PC program for fitting calibration curves in chromosome aberration dosimetry. Radiat Protect Dosimetry. 124:115–123.
12. Di Giorgio M, Barquinero JF, Vallerga MB, Radl A, Taja MR, Seoane A, De Luca J, Stuck Oliveira M, Valdivia P, Garcıa Lima O, Lamadrid A, Gonzalez Mesa J, Romero Aguilera I, Mandina Cardoso T, Guerrero Carvajal YC, Arceo Maldonado C, Espinoza ME, Martınez-Lopez W, Mendez-Acuna L, Di Tomaso MV, Roy L, Lindholm C, Romm H, Guclu I and Lloyd D. 2011. Biological dosimetry intercomparison exercise: an evaluation of triage and routine mode results by robust methods. Radiat Res. 175:638–649.
13. Esseen CG. 1942. On the Liapunoff limit of error in the theory of probability. Arkiv for Matematik Astronomi och Fysik. A28 (9):1-19.
14. Garcia O, Di Giorgio M, Vallerga MB, Radl A, Taja MR, Seoane A, De Luca J, Stuck Oliveira M, Valdivia P, Lamadrid AI, Gonzalez JE, Romero I, Mandina T, Pantelias G, Terzoudi G, Guerrero-Carbajal C, Arceo Maldonado C, Espinoza M, Oliveros N, Martınez-Lopez W, Di Tomaso MV, Mendez-Acuna L, Puig R, Roy L and Barquinero JF. 2013. Interlaboratory comparison of dicentric chromosome assay using electronically transmitted images. Radiat Prot Dosimetry. 154(1):18–25.
15. Gregoire E, Kulka U, Barrios L, Ainsbury E, Bassinet C, Fattibene P, Oestreicher U, Pantelias, G, Terzoudi G, Trompier F, Voisin P, Vral A, Wojcik A and Roy L. 2017 Jan. The harmonization process to set up and maintain an operational biological dosimetry and physical retrospective dosimetry network: QA QM applied to the RENEB network. Int J Rad Biol. 93(1): 81-86.
16. International Atomic Energy Agency (IAEA). 2001. Manual 405 Vienna: IAEA
17. International Atomic Energy Agency (IAEA). 2011. Cytogenetic dosimetry: applications in preparedness for and response to radiation emergencies. Vienna: IAEA.

18. International Organization for Standardization (ISO) 13528: 2015. Statistical methods for use in proficiency testing by interlaboratory comparison.

19. International Organization for Standardization (ISO) 5725. 1998. Precision of test methods-determination of repeatability & reproducibility for a standard test method by interlaboratory tests..

20. International Organization for Standardization (ISO). 19238. 2014. Radiation protection-performance criteria for service laboratories performing biological dosimetry by cytogenetics. Geneva: ISO.

21. Jaworska A, Ainsbury EA, Fattibene P, Lindholm C, Oestreicher U, Rothkamm K, Romm H, Thierens H, Trompier F, Voisin P, Vral A, Woda C and Wojcik A. 2015 Apr. Operational guidance for radiation emergency response organizations in Europe for using biodosimetric tools developed in EU MULTIBIODOSE project. Radiat Protect Dosimetry. 164:1–5.

22. Kulka U, Ainsbury, EA, Atkinson M, Barquinero JF, Barrios L, Beinke C, Bognar G, Cucu A, Darroudi F, Fattibene P, Gil O, Gregoire E, Hadjidekova V, Haghdoost, Herranz R, Jaworska A, Lindholm C, M'kacher R, Möertl S, Montoro A, Moquet J, Moreno M, Ogbazghi A, Oestreicher U, Palitti F, Pantelias G, Popescu I, Prieto MJ, Romm H, Rothkamm K, Sabatier L, Sommer S, Terzoudi G, Testa A, Thierens H, Trompier F, Turai I, Vandesickel V, Vaz P, Voisin P, Vral A, Ugletveit F, Woda C and Wojcik A. Realising the European network of biodosimetry (RENEB). 2012. Radiat. Prot. Dosimetry. 151(4): 621-625.

23. Kulka U, Ainsbury, EA, Atkinson M, Barnard S, Smith R, Barquinero JF, Barrios L, Bassinet C, Beinke C, Cucu A, Darroudi F, Fattibene P, Bortolin F, Della Monaca S, Gil O, Gregoire E, Hadjidekova V, Haghdoost S, Hatzi V, Hempel W, Herranz R, Jaworska A, Lindholm C, Lumniczky K, M'kacher R, Möertl S, Montoro A, Moquet J, Moreno M, Noditi M, Ogbazghi A, Oestreicher U, Palitti F, Pantelias G, Popescu I, Prieto MJ, Roch-Lefévre S, Roessler U, Romm H, Rothkamm K, Sabatier L, Sebastia N, Sommer S, Terzoudi G, Testa A, Thierens H, Trompier F, Turai I, Vandevoorde C, Vaz P, Voisin P, Vral A, Ugletveit F, Wieser A, Woda C and Wojcik A. Realising the European network of biodosimetry: RENEB-status quo. 2015. Radiat. Prot. Dosimetry. 164:42–45.

24. Kulka U, Abend M, Ainsbury E, Badie C, Barquinero JF, Barrios L, Beinke C, Bortolin E, Cucu A, De Amicis A, Dominguez I, Fattibene P, Frovig AM, Gregoire E, Guogyte K, Hadjidekova V, Jaworska A, Kriehuber R, Lindholm C, Lloyd D, Lumniczky K, Lyng F, Meschini R, Mörtl S, Della

Monaca S, Monteiro Gil O, Montoro A, Moquet J, Moreno M, Oestreicher U, Palitti F, Pantelias G, Patrono C, Piqueret-Stephan L, Port M, Prieto MJ, Quintens R, Ricoul M, Romm H, Roy L, Sáfrány G, Sabatier L, Sebastià N, Sommer S, Terzoudi G, Testa A, Thierens H, Turai I, Trompier F, Valente M, Vaz P, Voisin P, Vral A, Woda C, Zafiropoulos D and Wojcik A. 2017. RENEB – Running the European Network of biological dosimetry and physical retrospective dosimetry. Int J Rad Biol. 93(1):2-14.

25. Liu JX, Pan Y, Ruan JL, Piao C, Su X. 2016. Intercomparison in cytogenetic dosimetry among 22 laboratories in China. Genome Integr. 7:6.
26. Lloyd DC, Edwards AA, Moquet JE, Guerrero-Carbajal YC. 2000. The role of cytogenetics in early triage of radiation casualties. Appl Radiat Isot. 52:1107-1112.
27. Merkle W. 1983. Statistical methods in regression and calibration analysis of chromosome aberration data. Radiat Environ Biophys. 21(3):217–233.
28. Miller SM, Ferrarotto CL, Vlahovich S, Wilkins RC, Boreham DR and Dolling J A. 2007. Canadian cytogenetic emergency network (CEN) for biological dosimetry following radiological/nuclear accidents. Int J Rad Biol. 83:471–477.
29. Oestreicher U, Samaga D, Ainsbury E, Antunes AC, Baeyens A, Barrios L, Beinke C, Beukes P, Blakely WF, Cucu A, De Amicis A, Depuydt J, De Sanctis S, Di Giorgio M, Dobos K, Dominguez I, Ngoc Duy P, Espinoza ME, Flegal FN, Figel M, Garcia O, Monteiro Gil O, Gregoire E, Guerrero-Carbajal C, Güçlü İ, Hadjidekova V, Hande P, Kulka U, Lemon J, Lindholm C, Lista F, Lumniczky K, Martinez-Lopez W, Maznyk N, Meschini R, M'kacher R, Montoro A, Moquet J, Moreno M, Noditi M, Pajic J, Radl A, Ricoul M, Romm H, Roy L, Sabatier L, Sebastià N, Slabbert J, Sommer S, Stuck Oliveira M, Subramanian U, Suto Y, Que T, Testa A, Terzoudi G, Vral A, Wilkins R, Yanti LY, Zafiropoulos D and Wojcik A. 2017. RENEB intercomparisons applying the conventional Dicentric Chromosome Assay (DCA). Int J Rad Biol. 93(1):20-29.
30. Pan Y, Ruan J, Gao G, Wu L, Piao C, and Liu J. 2019 Jan-Mar. Laboratory intercomparison of cytogenetic dosimetry among 38 laboratories in china. Dose-Response. 17(1):1-7.
31. Romm H, Wilkins RC, Coleman CN, Lillis-Hearne PK, Pellmar TC, Livingston GK, Awa AA, Jenkins MS, Yoshida MA, Oestreicher U and Prasanna PGS. 2011. Biological dosimetry by the triage dicentric chromosome assay: potential implications for treatment of acute radiation syndrome in radiological mass casualties. Radiat Res. 175:397–404.

32. Romm H, Ainsbury E, Bajinskis A, Barnard S, Barquinero JF, Beinke C, Puig-Casanovas R, Deperas-Kaminska M, Gregoire E, Kulka U, Oestreicher U, Lindholm C, Moquet J, Rothkamm K, Sommer S, Thierens H, Vral A, Vandersickel V, Wojcik A. 2014a. Web-based scoring of the dicentric assay, a collaborative biodosimetric scoring strategy for population triage in large scale radiation accidents. Radiat Environ Biophys. 53(2):241-254.

33. Romm H, Ainsbury E, Barnard S, Barrios L, Barquinero JF, Beinke C, Deperas M, Gregoire E, Koivistoinen A, Lindholm C, Moquet J, Oestreicher U, Puig R, Rothkamm K, Sommer S, Thierens H, Vandersickel V, Vral A, Wojcik A. 2014b June. Validation of semi-automatic scoring of dicentric chromosomes after simulation of three different irradiation scenarios, Health Phys. 106(6):764-71.

34. Romm H, Beinke C, Garcia O, Di Giorgio M, Gregoire E, Livingston G, Lloyd D, Martinez-Lopez W, Moquet JE, Sugarman SL, Wilkins RC and Ainsbury EA. 2016. A new cytogenetic biodosimetry image repository for the dicentric assay. Radiat Prot Dosimetry. 172(1-3): 192–200

35. Roy L, Buard V, Delbos M, Durand V, Paillole N, Gregoire E and Voisin P. 2004. International intercomparison for criticality dosimetry: the case of biological dosimetry. Radiat Prot Dosimetry. 110(1-4):471–476.

36. Shaffer JP. 1995. Multiple hypothesis testing. Annual Review of Psychology. 46:561-584.

37. Trompier F, Baumann M, Barrios L, Gregoire E, Abend M, Ainsbury E, Barnard S, Barquinero JF, Bautista JA, Brzozowska B, Perez-Calatayud J, De Angelis C, Domínguez I, Hadjidekova V, Kulka U, Mateos JC, Meschini R, Monteiro Gil O, Moquet J, Oestreicher U, Montoro Pastor A, Quintens R, Sebastià N, Sommer S, Stoyanov O, Thierens H, Terzoudi G, Villaescusa JI, Vral A, Wojcik A, Zafiropoulos D and Roy L. 2017 Jan. Investigation of the influence of calibration practices on cytogenetic laboratory performance for dose estimation. Int J Rad Biol. 93(1):118-126.

38. Vaurijoux A, Gruel G, Gregoire E, Roch-Lefevre S, Voisin Pa, Martin C, Voisin Ph, Roy L, Barquinero JF. 2015. Automatic dicentric scoring a real option to be used in biological dosimetry. Rad Emerg Med. 4:16-21.

39. Voisin P. Standards in biological dosimetry: A requirement to perform an appropriate dose assessment. 2015. Mutat Res. 793:115–122.

40. Wilkins R, Romm H, Kao TC, Awa AA, Yoshida MA, Livingston GK, Jenkins MS, Oestreicher U, Pellmar TC, and Prasanna PGS. 2008. Interlaboratory comparison of the dicentric chromosome assay for radiation biodosimetry in mass casualty events. Radiat Res. 169(5):551-560.

41. Wilkins RC, Romm H, Oestreicher U, Marro L, Yoshida M A, Suto Y, and Prasanna PGS. 2011 Sept. Biological dosimetry by the triage dicentric chromosome assay – further validation of international networking. Radiat Meas. 46(9): 923–928.

42. Wojcik A, Lloyd D, Romm H and Roy L, Biological dosimetry for triage of casualties in a large-scale radiological emergency: capacity of the EU member states. 2010. Radiat Prot Dosimetry. 138(4):397–401.

43. Yoshida MA, Hayata I, Tateno H, Tanaka K, Sonta S, Kodama S, Kodama Y and Sasaki MS. 2007. The chromosome network for biodosimetry in Japan. Radiat Meas. 42:1125–1127.

Figures and Tables Legends

**Table 1. Reported scoring values from each laboratory; dicentric (dic) and dicentric plus ring (dic+r).** In italics, reported values that did not include the cell distribution of aberrations. L1 to L20 are RENEB members and L21 to L38 are non-RENEB group. *L2 and L2b represent different kinds of staining (Giemsa (L2) or FISH coupled with pan-telomeric and pan-centromeric probes (L2b)). **L31 and L31b represent chromosomal aberration scoring (dicentrics (L31) and dicentrics plus rings (L31b)).

**Figure 1: Dicentric frequencies per metaphase for sample A (A), sample B (B) and sample C (C) from each of the participant laboratories.** Triangles represent the dicentric frequency per metaphase obtained by each laboratory for sample A (A), sample B (B) and sample C (C). The solid black line is the robust mean when laboratories reported the dicentric frequency per metaphase. Dashed lines mark the 95% confidence interval of the robust mean. Z-and U-scores were calculated with a Benjamini-Hochberg adjustment. Stars denote an unsatisfactory U-score and circles denote an unsatisfactory Z score. RENEB laboratories are within the grey rectangle.

**Table 2. Calibration curve coefficients of the participant laboratories. NA**: Not Available. These laboratories have no dose-effect curve. Some laboratories did not include standard deviations for the coefficients (± NA). ***L2 and L2b** represent different kinds of staining (Giemsa (L2) or FISH coupled with pan-telomeric and pan-centromeric probes (L2b)). ****L4/L5 and L4b/L5b** represent calibration curves based on different irradiation sources (gamma- (L4 and L5) and x-rays (L4b and L5b)). *****L31 and L31b** represent chromosomal aberration scoring (dicentrics (L31) and dicentrics plus rings (L31b)). L1 to L20 are RENEB members and L21 to L38 are non-RENEB group.

**Table 3: Source and dose rate used by the laboratories for their calibration curve.**

* : L2 used Giemsa for dicentric chromosomes and centric rings staining, and L2b used TC-FISH for dicentric chromosomes and centric rings staining. ** : L31 scored only dicentric chromosomes to

build its curve and L31b scored dicentric chromosomes and centric rings to build its curve. L1 to L20 are RENEB members and L21 to L38 are non-RENEB group. **NP** : Not Provided

**Figure 2: Dose estimations sent by the participant laboratories for samples A, B, C.** Solid circles represent the dose estimation based on the total number of metaphases analyzed. Some laboratories sent two estimated doses for each sample, which are represented by open triangles. Error bars correspond to the reported 95% confidence interval. The horizontal line represents the delivered physical dose to the blood.

**Figure 3: Re-calculated doses by the IRSN based on Merckle's approach and using each laboratory's own curve coefficients for sample A (3A), sample B (3B) and sample C (3C).** Diamonds represent the average dose obtained and error bars correspond to the 95% confidence interval of the estimated dose. Values considered as unsatisfactory by the U-test are indicated with a star. Unsatisfactory results by the Z-test are indicated with a circle. RENEB laboratories are within the grey rectangle.

**Table 4: Laboratory ranking by Z-Score (A) and by U-Score (B). a** : CA Frequency : Frequency of chromosomal aberrations (dicentric chromosomes per cell).

**Table 5: Comparison of laboratory rankings between the Z-score obtained for dicentric frequency per metaphase and the Z-score obtained for assessed dose.** Rectangles show the laboratories whose rank changes the most between frequency and dose. L9 is highlighted by a solid line rectangle and L18 is highlighted by a dashed line rectangle. L36, L37 and L38 are not present in the dose column since they did not provide dose estimations. L4b, L5b and L31b are present only in the dose column because the dicentric frequencies are similar within the same laboratory (L4/L4b ; L5/L5b ; L31/L31b). **NA**: Not Available: The Z-score for L2b was not calculated because the staining technique (TC-FISH) was different from the rest (GIEMSA staining) and thus could not be compared using this test.

**Figure 4: Calibration curves of the Inter-Laboratory Comparison participants.** The horizontal line represents the frequency of 0.5 dicentric chromosomes or dicentrics + centric rings per metaphase

and the grey vertical lines indicate the mean estimated dose obtained with the two most distant curves. The dashed curve, indicated by an arrow, is that of the IAEA manual (IAEA 2011).

**Table 6: Comparison of robust values among laboratory categories.** x, s and CV correspond respectively to the calculated robust mean, robust standard deviation and robust coefficient of variation calculated.

| Laboratory | SAMPLE | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | | | B | | | C | | |
| | cells | dic | dic+r | cells | dic | dic+r | cells | dic | dic+r |
| **L1** | 500 | 106 | - | 500 | 0 | - | 500 | 8 | - |
| **L2*** | 501 | 104 | 120 | 511 | 0 | 0 | 500 | 14 | 19 |
| **L2b*** | 517 | 158 | 185 | 570 | 1 | 1 | 656 | 27 | 35 |
| **L3** | 300 | 98 | 104 | 300 | 0 | 0 | 300 | 13 | 14 |
| **L4** | 500 | 127 | - | 500 | 0 | - | 500 | 8 | - |
| **L5** | 500 | 122 | - | 500 | 3 | - | 525 | 9 | - |
| **L6** | 473 | 93 | 104 | 510 | 1 | 1 | 512 | 18 | 19 |
| **L7** | 15 | 3 | - | 38 | 0 | - | 12 | 1 | - |
| **L8** | 500 | 116 | - | 500 | 2 | - | 500 | 14 | - |
| **L9** | 500 | 113 | 116 | 500 | 0 | 0 | 500 | 5 | 6 |
| **L10** | 520 | 116 | 124 | 500 | 0 | 0 | 531 | 8 | 9 |
| **L11** | 250 | 31 | 33 | 250 | 0 | 0 | 250 | 7 | 7 |
| **L12** | 500 | 93 | - | 500 | 0 | - | 500 | 12 | - |
| **L13** | 259 | 51 | 57 | 344 | 0 | 0 | 298 | 7 | 7 |
| **L14** | 500 | 90 | - | 500 | 5 | - | 500 | 9 | - |
| **L15** | 500 | 118 | - | 500 | 1 | - | 500 | 14 | - |
| **L16** | 500 | 98 | - | 500 | 1 | - | 500 | 13 | - |
| **L17** | 500 | 87 | 89 | 500 | 0 | 0 | 500 | 24 | 24 |
| **L18** | 540 | 167 | - | 500 | 1 | - | 526 | 19 | - |
| **L19** | 474 | 79 | 88 | 500 | 0 | 0 | 477 | 10 | 11 |
| **L20** | 500 | 143 | - | 500 | 4 | - | 500 | 17 | - |
| **L21** | 421 | 100 | 106 | 500 | 3 | 3 | 603 | 14 | 15 |
| **L22** | 500 | 151 | - | 500 | 0 | - | 500 | 20 | - |
| **L23** | 427 | 100 | 100 | 500 | 1 | 1 | 500 | 11 | 11 |
| **L24** | 500 | 101 | 104 | 500 | 1 | 1 | 500 | 10 | 10 |
| **L25** | 500 | 105 | - | 500 | 0 | - | 500 | 8 | - |
| **L26** | 500 | 121 | 121 | 500 | 0 | 0 | 500 | 11 | 11 |
| **L27** | 500 | 104 | - | 500 | 1 | - | 500 | 11 | - |
| **L28** | 500 | 160 | - | 500 | 0 | - | 500 | 20 | - |
| **L29** | 500 | 128 | - | 500 | 5 | - | 500 | 15 | - |
| **L30** | 500 | 86 | 86 | 500 | 1 | 1 | 500 | 14 | 15 |
| **L31**** | 500 | 129 | 147 | 500 | 1 | - | 500 | 10 | 10 |
| **L31b**** | 500 | 129 | - | 500 | 1 | 1 | 500 | 10 | - |
| **L32** | 500 | 71 | 75 | 500 | 0 | 0 | 500 | 5 | 5 |
| **L33** | 256 | 59 | - | 500 | 0 | - | 500 | 15 | - |
| **L34** | 500 | 98 | - | 500 | 1 | - | 500 | 12 | - |
| **L35** | 500 | 172 | - | 500 | 0 | - | 500 | 12 | - |
| **L36** | *500* | *53* | *57* | *500* | *1* | *1* | *500* | *0* | *0* |
| **L37** | *500* | *56* | *63* | *500* | *1* | *1* | *500* | *2* | *2* |
| **L38** | *200* | *20* | *33* | *500* | *3* | *4* | *200* | *12* | *12* |

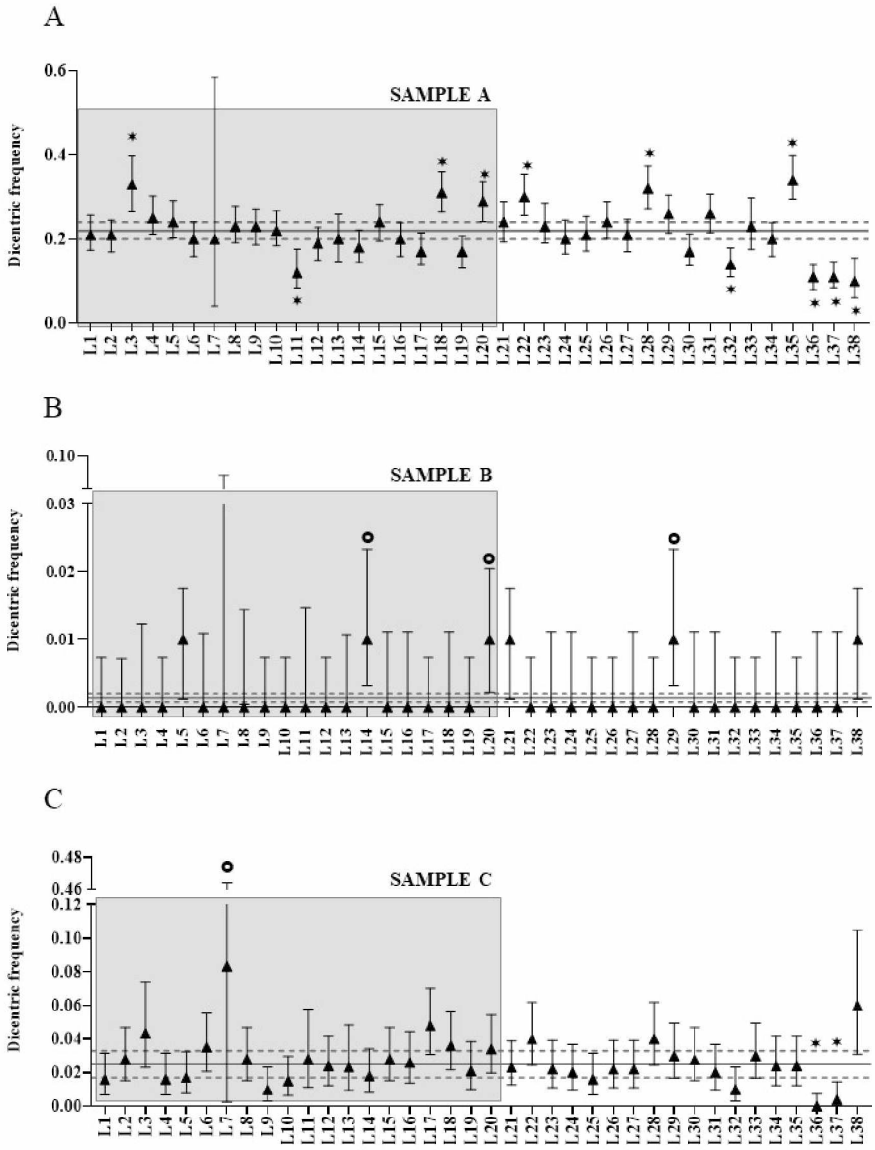



Figure 1: Dicentric frequencies per metaphase for sample A (A), sample B (B) and sample C (C) from each of the participant laboratories. Triangles represent the dicentric frequency per metaphase obtained by each laboratory for sample A (A), sample B (B) and sample C (C). The solid black line is the robust mean when laboratories reported the dicentric frequency per metaphase. Dashed lines mark the 95% confidence interval of the robust mean. Z-and U-scores were calculated with a Benjamini-Hochberg adjustment. Stars denote an unsatisfactory U-score and circles denote an unsatisfactory Z score. RENEB laboratories are within the grey rectangle.

190x254mm (96 x 96 DPI)

| Laboratory | α ± SE: | | | β ± SE: | | | C ± SE: | | |
|---|---|---|---|---|---|---|---|---|---|
| **L1** | 0.0187 | ± | 0.0047 | 0.0527 | ± | 0.0039 | 0.0002 | ± | 0.0001 |
| **L2*** | 0.0573 | ± | 0.0090 | 0.0778 | ± | 0.0037 | 0.0012 | ± | 0.0006 |
| **L2b*** | 0.0956 | ± | 0.0248 | 0.1010 | ± | 0.0089 | 0.0014 | ± | 0.0017 |
| **L3** | 0.0057 | ± | 0.0053 | 0.0817 | ± | 0.0051 | 0.0012 | ± | 0.0006 |
| **L4**** | 0.0120 | ± | 0.0030 | 0.0560 | ± | 0.0020 | 0.0010 | ± | 0.0040 |
| **L4b**** | 0.0350 | ± | 0.0030 | 0.0640 | ± | 0.0020 | 0.0010 | ± | 0.0040 |
| **L5**** | 0.0197 | ± | 0.0064 | 0.0597 | ± | 0.0037 | 0.0012 | ± | 0.0010 |
| **L5b**** | 0.0537 | ± | 0.0123 | 0.0626 | ± | 0.0218 | 0.0006 | ± | 0.0017 |
| **L6** | 0.0399 | ± | 0.0061 | 0.0485 | ± | 0.0029 | 0.0010 | ± | 0.0003 |
| **L7** | 0.0105 | ± | 0.0035 | 0.0480 | ± | 0.0019 | 0.0011 | ± | 0.0006 |
| **L8** | 0.0413 | ± | 0.0058 | 0.0444 | ± | 0.0033 | 0.0007 | ± | 0.0060 |
| **L9** | 0.0283 | ± | 0.0056 | 0.0255 | ± | 0.0030 | 0.0008 | ± | 0.0005 |
| **L10** | 0.0193 | ± | 0.0024 | 0.0612 | ± | 0.0036 | 0.0004 | ± | 0.0002 |
| **L11** | 0.0117 | ± | 0.0020 | 0.0456 | ± | 0.0064 | 0.0043 | ± | 0.0017 |
| **L12** | 0.0142 | ± | 0.0044 | 0.0759 | ± | 0.0027 | 0.0005 | ± | 0.0005 |
| **L13** | 0.0101 | ± | 0.0051 | 0.0720 | ± | 0.0043 | 0.0006 | ± | 0.0004 |
| **L14** | 0.0690 | ± | 0.0230 | 0.0310 | ± | 0.0110 | 0.0000 | ± | 0.0000 |
| **L15** | 0.0229 | ± | 0.0049 | 0.0622 | ± | 0.0039 | 0.0081 | ± | 0.0030 |
| **L16** | 0.0419 | ± | 0.0080 | 0.0529 | ± | 0.0018 | 0.0010 | ± | 0.0063 |
| **L17** | 0.0283 | ± | NA | 0.0718 | ± | NA | 0.0005 | ± | NA |
| **L18** | 0.0612 | ± | 0.0097 | 0.0650 | ± | 0.0052 | 0.0008 | ± | 0.0004 |
| **L19** | 0.0073 | ± | 0.0194 | 0.0668 | ± | 0.0046 | 0.0008 | ± | 0.0000 |
| **L20** | 0.0322 | ± | 0.0123 | 0.0459 | ± | 0.0069 | 0.0028 | ± | 0.0015 |
| **L21** | 0.0210 | ± | 0.0052 | 0.0631 | ± | 0.0040 | 0.0013 | ± | 0.0005 |
| **L22** | 0.0210 | ± | 0.0052 | 0.0631 | ± | 0.0040 | 0.0013 | ± | 0.0005 |
| **L23** | 0.0355 | ± | 0.0041 | 0.0644 | ± | 0.0027 | 0.0011 | ± | 0.0001 |
| **L24** | 0.0120 | ± | 0.0080 | 0.0510 | ± | 0.0030 | 0.0010 | ± | 0.0001 |
| **L25** | 0.0208 | ± | 0.0045 | 0.0710 | ± | 0.0033 | 0.0002 | ± | 0.0002 |
| **L26** | 0.0331 | ± | 0.0077 | 0.0359 | ± | 0.0036 | 0.0012 | ± | 0.0006 |
| **L27** | 0.0419 | ± | 0.0017 | 0.0890 | ± | 0.0047 | 0.0027 | ± | 0.0008 |
| **L28** | 0.0813 | ± | 0.0046 | 0.0824 | ± | 0.0021 | 0.0025 | ± | 0.0004 |
| **L29** | 0.0356 | ± | 0.0096 | 0.0779 | ± | 0.0065 | 0.0021 | ± | 0.0014 |
| **L30** | 0.0682 | ± | 0.0038 | 0.0344 | ± | 0.0066 | 0.0013 | ± | 0.0008 |
| **L31***** | 0.0416 | ± | 0.0080 | 0.0585 | ± | 0.0073 | 0.0009 | ± | 0.0002 |
| **L31b***** | 0.0305 | ± | 0.0079 | 0.0624 | ± | 0.0113 | 0.0009 | ± | 0.0005 |
| **L32** | 0.0209 | ± | 0.0057 | 0.0711 | ± | 0.0025 | 0.0005 | ± | 0.0002 |
| **L33** | 0.0421 | ± | 0.0042 | 0.0602 | ± | 0.0022 | 0.0009 | ± | 0.0003 |
| **L34** | 0.0419 | ± | 0.0080 | 0.0529 | ± | 0.0018 | 0.0010 | ± | 0.0063 |
| **L35** | 0.0313 | ± | NA | 0.0537 | ± | NA | -0.0078 | ± | NA |
| **L36** | NA | | NA | NA | | NA | NA | | NA |
| **L37** | NA | | NA | NA | | NA | NA | | NA |
| **L38** | NA | | NA | NA | | NA | NA | | NA |

| Laboratory | Source | Dose Rate (Gy/min) |
|---|---|---|
| **L1** | 137 Cs | 0.42 |
| ***L2** | 60 Cobalt | 0.5 |
| ***L2b** | 60 Cobalt | 0.5 |
| **L3** | 60 Cobalt | 269mGy/7min: 0.04 |
| **L4** | 60 Cobalt | 0.35 |
| **L4b** | 243 kV X Rays | 0.35 |
| **L5** | 60 Cobalt | 0.3 |
| **L5b** | 200 kV X Rays | 1.271 |
| **L6** | 60 Cobalt | 0.5 |
| **L7** | 60 Cobalt | 0.18-0.13 |
| **L8** | 60 Cobalt | 0.5774 |
| **L9** | 137 Cs | 0.87 |
| **L10** | 60 Cobalt | 0.5 |
| **L11** | 60 Cobalt | <0.5 |
| **L12** | 60 Cobalt | 0.5 |
| **L13** | 60 Cobalt | 0.24 |
| **L14** | 60 Cobalt | 1 |
| **L15** | 60 Cobalt | 1 |
| **L16** | 60 Cobalt | 1Gy/5min: 0.2 |
| **L17** | NP | NP |
| **L18** | 60 Cobalt | 0.3 |
| **L19** | 60 Cobalt | 0.5 |
| **L20** | 6 MeV X Rays | 2.5 |
| **L21** | 60 Cobalt | 0.5 |
| **L22** | 60 Cobalt | ND |
| **L23** | 60 Cobalt | 0.5 |
| **L24** | 60 Cobalt | 0.7 |
| **L25** | 60 Cobalt | 0.275 |
| **L26** | 60 Cobalt | 0.66 |
| **L27** | 250 kVp X-rays | 0.37 |
| **L28** | 250 kVp X-rays | 0.37 |
| **L29** | 137 Cs | 0.94 |
| **L30** | 250 kVp X-rays | 0.37 |
| ****L31** | 250 kVp X-rays | 0.37 |
| ****L31b** | 250 kVp X-rays | 0.37 |
| **L32** | 60 Cobalt | 0.457 |
| **L33** | 60 Cobalt | 0.16 |
| **L34** | 60 Cobalt | 0.2 |
| **L35** | 137 Cs | 1.16 |
| **L36** | NP | NP |
| **L37** | NP | NP |
| **L38** | NP | NP |

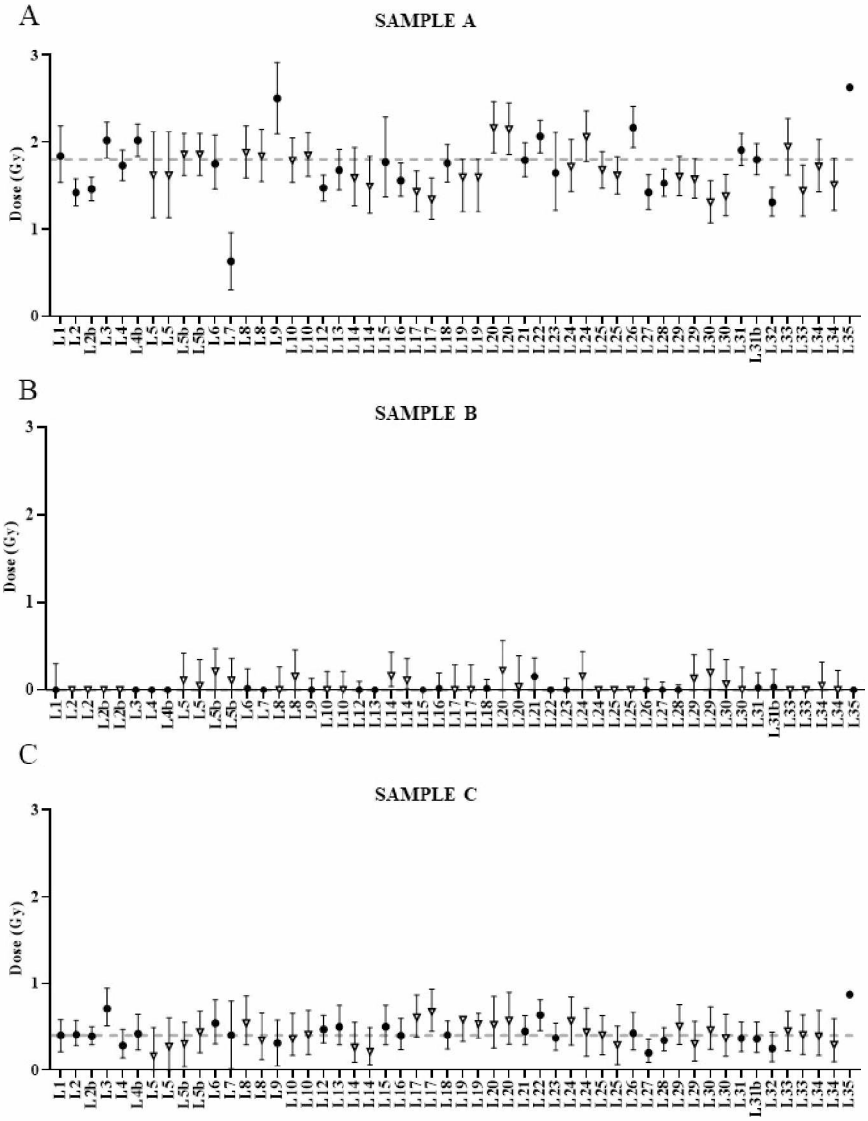



Figure 2: Dose estimations sent by the participant laboratories for samples A, B, C. Solid circles represent the dose estimation based on the total number of metaphases analyzed. Some laboratories sent two estimated doses for each sample, which are represented by open triangles. Error bars correspond to the reported 95% confidence interval. The horizontal line represents the delivered physical dose to the blood.

190x254mm (96 x 96 DPI)

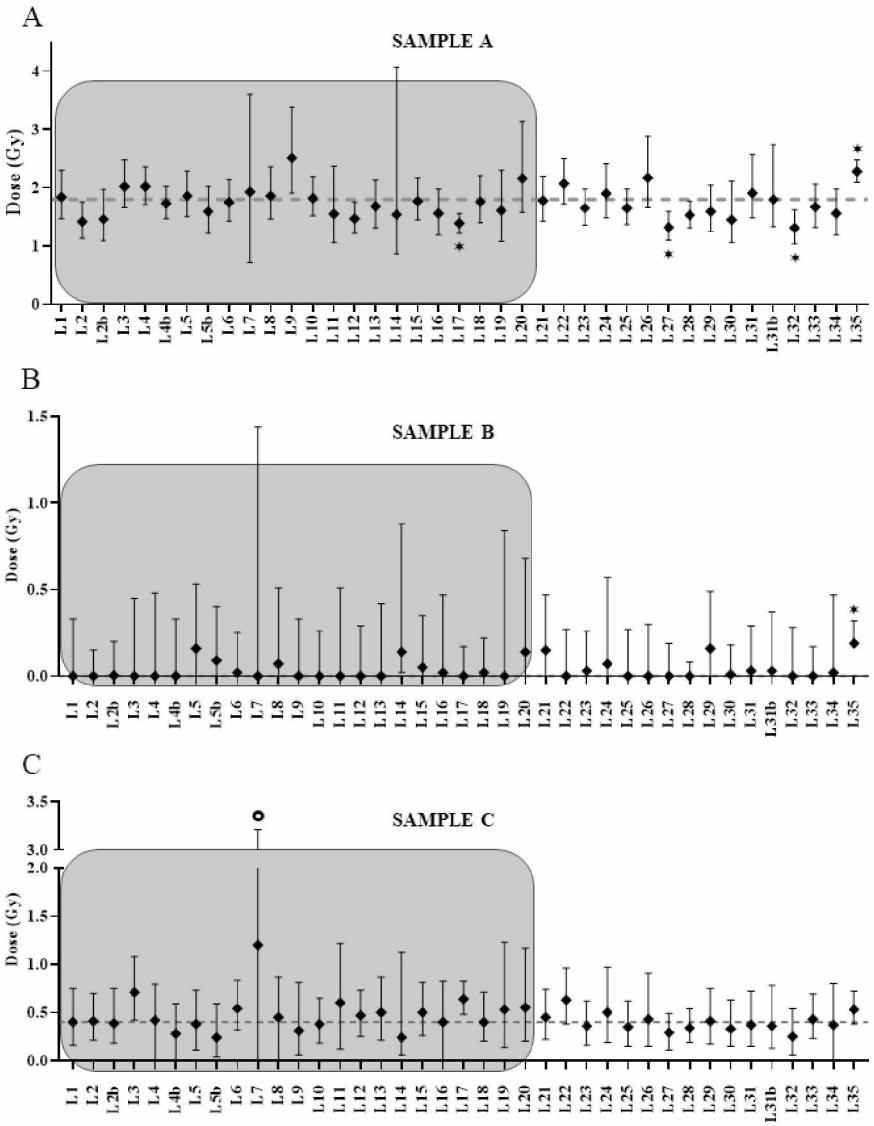



Figure 3: Re-calculated doses by the IRSN based on Merckle's approach and using each laboratory's own curve coefficients for sample A (3A), sample B (3B) and sample C (3C). Diamonds represent the average dose obtained and error bars correspond to the 95% confidence interval of the estimated dose. Values considered as unsatisfactory by the U-test are indicated with a star. Unsatisfactory results by the Z-test are indicated with a circle. RENEB laboratories are within the grey rectangle.

190x254mm (96 x 96 DPI)

A

| Rank | Laboratory | Analyzed Metaphases | CA Frequency[a] | Z Score |
|---|---|---|---|---|
| **1** | L1 | 500 | 0.212 | 0.083 |
| **2** | L10 | 520 | 0.223 | 0.104 |
| **3** | L25 | 500 | 0.210 | 0.117 |
| **4** | L27 | 500 | 0.208 | 0.150 |
| **5** | L9 | 500 | 0.226 | 0.154 |
| **6** | L2 | 501 | 0.208 | 0.157 |
| **7** | L33 | 256 | 0.230 | 0.229 |
| **8** | L24 | 500 | 0.202 | 0.252 |
| **9** | L8 | 500 | 0.232 | 0.255 |
| **10** | L7 | 15 | 0.200 | 0.286 |
| **11** | L23 | 427 | 0.234 | 0.292 |
| **12** | L15 | 500 | 0.236 | 0.323 |
| **13** | L13 | 259 | 0.197 | 0.338 |
| **14** | L6 | 473 | 0.197 | 0.343 |
| **15** | L21 | 421 | 0.238 | 0.349 |
| **16** | L16 | 500 | 0.196 | 0.353 |
| **17** | L34 | 500 | 0.196 | 0.353 |
| **18** | L26 | 500 | 0.242 | 0.424 |
| **19** | L5 | 500 | 0.244 | 0.458 |
| **20** | L12 | 500 | 0.186 | 0.522 |
| **21** | L14 | 500 | 0.180 | 0.624 |
| **22** | L4 | 500 | 0.254 | 0.627 |
| **23** | L29 | 500 | 0.256 | 0.661 |
| **24** | L31 | 500 | 0.258 | 0.695 |
| **25** | L17 | 500 | 0.174 | 0.725 |
| **26** | L30 | 500 | 0.172 | 0.759 |
| **27** | L19 | 474 | 0.167 | 0.849 |
| **28** | L20 | 500 | 0.286 | 1.168 |
| **29** | L32 | 500 | 0.142 | 1.266 |
| **30** | L22 | 500 | 0.302 | 1.438 |
| **31** | L18 | 540 | 0.309 | 1.561 |
| **32** | L11 | 250 | 0.124 | 1.570 |
| **33** | L28 | 500 | 0.320 | 1.743 |
| **34** | L37 | 500 | 0.112 | 1.773 |
| **35** | L3 | 300 | 0.327 | 1.855 |
| **36** | L36 | 500 | 0.106 | 1.874 |
| **37** | L38 | 200 | 0.100 | 1.976 |
| **38** | L35 | 500 | 0.344 | 2.148 |

B

| Rank | Laboratory | Analyzed Metaphases | CA Frequency[a] | U-Score |
|---|---|---|---|---|
| **1** | L7 | 15 | 0.200 | 0.124 |
| **2** | L1 | 500 | 0.212 | 0.206 |
| **3** | L10 | 520 | 0.223 | 0.258 |
| **4** | L25 | 500 | 0.210 | 0.291 |
| **5** | L9 | 500 | 0.226 | 0.373 |
| **6** | L27 | 500 | 0.208 | 0.398 |
| **7** | L33 | 256 | 0.230 | 0.415 |
| **8** | L2 | 501 | 0.208 | 0.417 |
| **9** | L8 | 500 | 0.232 | 0.613 |
| **10** | L24 | 500 | 0.202 | 0.637 |
| **11** | L23 | 427 | 0.234 | 0.656 |
| **12** | L13 | 259 | 0.197 | 0.657 |
| **13** | L15 | 500 | 0.236 | 0.771 |
| **14** | L21 | 421 | 0.238 | 0.774 |
| **15** | L6 | 473 | 0.197 | 0.857 |
| **16** | L16 | 500 | 0.196 | 0.903 |
| **17** | L34 | 500 | 0.196 | 0.903 |
| **18** | L26 | 500 | 0.242 | 1.003 |
| **19** | L5 | 500 | 0.244 | 1.079 |
| **20** | L12 | 500 | 0.186 | 1.360 |
| **21** | L4 | 500 | 0.254 | 1.455 |
| **22** | L29 | 500 | 0.256 | 1.529 |
| **23** | L31 | 500 | 0.258 | 1.603 |
| **24** | L14 | 500 | 0.180 | 1.643 |
| **25** | L17 | 500 | 0.174 | 1.934 |
| **26** | L30 | 500 | 0.172 | 2.032 |
| **27** | L19 | 474 | 0.167 | 2.253 |
| **28** | L20 | 500 | 0.286 | 2.589 |
| **29** | L3 | 300 | 0.327 | 3.114 |
| **30** | L22 | 500 | 0.302 | 3.120 |
| **31** | L18 | 540 | 0.309 | 3.463 |
| **32** | L11 | 250 | 0.124 | 3.604 |
| **33** | L32 | 500 | 0.142 | 3.616 |
| **34** | L28 | 500 | 0.320 | 3.694 |
| **35** | L35 | 500 | 0.344 | 4.422 |
| **36** | L38 | 200 | 0.100 | 4.474 |
| **37** | L37 | 500 | 0.112 | 5.457 |
| **38** | L36 | 500 | 0.106 | 5.864 |

| Ranks | Laboratory | Frequency (dicentric per metaphase) | Z Score | | Ranks | Laboratory | Dose (Gy) | Z-Score |
|---|---|---|---|---|---|---|---|---|
| **1** | L1 | 0.212 | 0.083 | | 1 | L31b | 1.8 | 0.000 |
| **2** | L10 | 0.223 | 0.104 | | 2 | L10 | 1.82 | 0.074 |
| **3** | L25 | 0.210 | 0.117 | | 3 | L21 | 1.78 | 0.074 |
| **4** | L27 | 0.208 | 0.150 | | 4 | L15 | 1.77 | 0.111 |
| **5** | L9 | 0.226 | 0.154 | | 5 | L1 | 1.84 | 0.148 |
| **6** | L2 | 0.208 | 0.157 | | 6 | L18 | 1.76 | 0.148 |
| **7** | L33 | 0.230 | 0.229 | | 7 | L6 | 1.75 | 0.184 |
| **8** | L24 | 0.202 | 0.252 | | 8 | L5 | 1.86 | 0.221 |
| **9** | L8 | 0.232 | 0.255 | | 9 | L8 | 1.86 | 0.221 |
| **10** | L7 | 0.200 | 0.286 | | 10 | L4b | 1.73 | 0.258 |
| **11** | L23 | 0.234 | 0.292 | | 11 | L24 | 1.9 | 0.369 |
| **12** | L15 | 0.236 | 0.323 | | 12 | L31 | 1.91 | 0.406 |
| **13** | L13 | 0.197 | 0.338 | | 13 | L13 | 1.68 | 0.443 |
| **14** | L6 | 0.197 | 0.343 | | 14 | L7 | 1.93 | 0.479 |
| **15** | L21 | 0.238 | 0.349 | | 15 | L33 | 1.67 | 0.479 |
| **16** | L16 | 0.196 | 0.353 | | 16 | L23 | 1.65 | 0.553 |
| **17** | L34 | 0.196 | 0.353 | | 17 | L25 | 1.65 | 0.553 |
| **18** | L26 | 0.242 | 0.424 | | 18 | L19 | 1.61 | 0.701 |
| **19** | L5 | 0.244 | 0.458 | | 19 | L5b | 1.59 | 0.775 |
| **20** | L12 | 0.186 | 0.522 | | 20 | L29 | 1.59 | 0.775 |
| **21** | L14 | 0.180 | 0.624 | | 21 | L3 | 2.02 | 0.811 |
| **22** | L4 | 0.254 | 0.627 | | 22 | L4 | 2.02 | 0.811 |
| **23** | L29 | 0.256 | 0.661 | | 23 | L16 | 1.56 | 0.885 |
| **24** | L31 | 0.258 | 0.695 | | 24 | L34 | 1.56 | 0.885 |
| **25** | L17 | 0.174 | 0.725 | | 25 | L11 | 1.55 | 0.922 |
| **26** | L30 | 0.172 | 0.759 | | 26 | L14 | 1.54 | 0.959 |
| **27** | L19 | 0.167 | 0.849 | | 27 | L22 | 2.07 | 0.996 |
| **28** | L20 | 0.286 | 1.168 | | 28 | L28 | 1.53 | 0.996 |
| **29** | L32 | 0.142 | 1.266 | | 29 | L12 | 1.47 | 1.217 |
| **30** | L22 | 0.302 | 1.438 | | 30 | L2b | 1.46 | 1.254 |
| **31** | L18 | 0.309 | 1.561 | | 31 | L30 | 1.45 | 1.291 |
| **32** | L11 | 0.124 | 1.570 | | 32 | L20 | 2.16 | 1.328 |
| **33** | L28 | 0.320 | 1.743 | | 33 | L26 | 2.17 | 1.365 |
| **34** | L37 | 0.112 | 1.773 | | 34 | L2 | 1.42 | 1.402 |
| **35** | L3 | 0.327 | 1.855 | | 35 | L17 | 1.39 | 1.512 |
| **36** | L36 | 0.106 | 1.874 | | 36 | L35 | 2.28 | 1.770 |
| **37** | L38 | 0.100 | 1.976 | | 37 | L27 | 1.32 | 1.770 |
| **38** | L35 | 0.344 | 2.148 | | 38 | L32 | 1.31 | 1.807 |
| | L2b | 0.305 | NA | | 39 | L9 | 2.51 | 2.619 |

**Mean Frequency : 0.22 dicentric per metaphase** **Dose : 1.74 Gy**

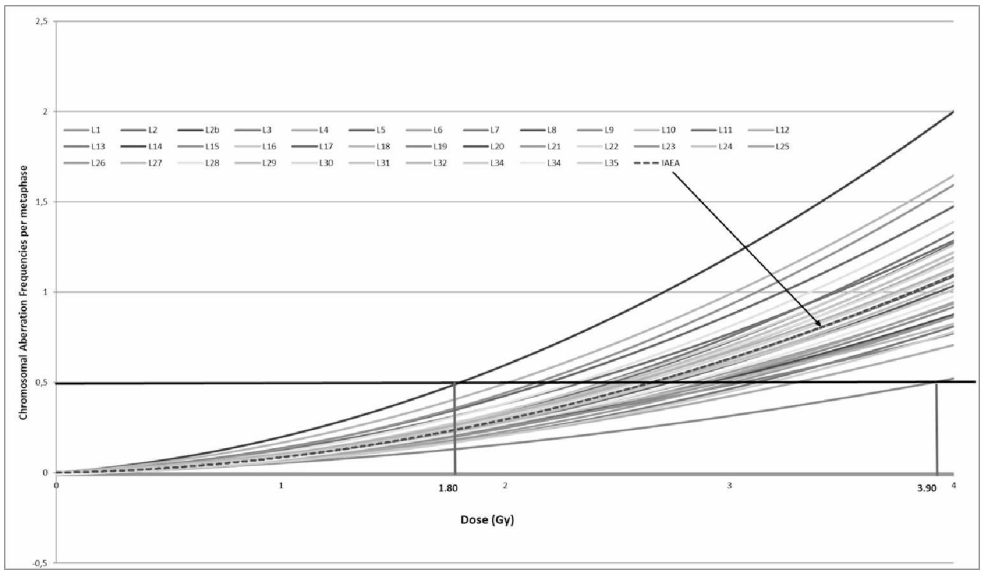



Figure 4: Calibration curves of the Inter-Laboratory Comparison participants. The horizontal line represents the frequency of 0.5 dicentric chromosomes or dicentrics + centric rings per metaphase and the grey vertical lines indicate the mean estimated dose obtained with the two most distant curves. The dashed curve, indicated by an arrow, is that of the IAEA manual (IAEA 2011).

338x190mm (96 x 96 DPI)

| Laboratories (Number) | x | s | CV |
|---|---|---|---|
| **ALL (38)** | 0.217 | 0.058 | 0.267 |
| **RENEB (20)** | 0.216 | 0.044 | 0.201 |
| **Non-RENEB (18)** | 0.214 | 0.078 | 0.365 |

# RENEB Inter-Laboratory Comparison 2017: limits and pitfalls of ILCs.

Eric Gregoire[1*], Joan Francesc Barquinero[2*], Gaetan Gruel[1*], Mohamedamine Benadjaoud[1*], Juan S. Martinez[1], Christina Beinke[3], Adayabalam Balajee[4], Philip Beukes[5], William F. Blakely[6], Inmaculada Dominguez[7], Pham Ngoc Duy[8], Octávia Monteiro Gil[9], Inci Güçlü[10], Kamile Guogyte[11], Savina Petrova Hadjidekova[12], Valeria Hadjidekova[13], Prakash Hande[14], Seongjae Jang[15], Katalin Lumniczky[16], Roberta Meschini[17], Mirta Milic[18], Alegria Montoro[19] Jayne Moquet[20], Mercedes Moreno[21], Farrah N Norton[22], Ursula Oestreicher[23], Jelena Pajic[24], Laure Sabatier[25], Sylwester Sommer[26], Antonella Testa[27], Georgia Terzoudi[28], Marco Valente[29], Perumal Venkatachalam[30], Anne Vral[31], Ruth C. Wilkins[32], Andrzej Wojcik[33], Demetre Zafiropoulos[34], Ulrike Kulka[23+].

* These authors contributed equally to this work
\+ Chair of RENEB e.V.

1. Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses, France
2. Universitat Autonoma de Barcelona, Barcelona, Spain
3. Bundeswehr Institute of Radiobiology affiliated to the University of Ulm, Munich, Germany
4. Oak Ridge Institute for Science and Education (ORISE), USA
5. NRF iThemba LABS, Cape Town, South Africa
6. Armed Forces Radiobiology Research Institute, Uniformed Service University of the Health Sciences, Bethesda, USA
7. University of Sevilla, Sevilla, Spain
8. Center of Biotechnology, Nuclear Research Institute, Nuclear Research Institute, Dalat city, Vietnam
9. Centro de Ciências e Tecnologias Nucleares, Instituto Superior Técnico, Universidade de Lisboa, Bobadela-LRS , Portugal
10. Turkish Atomic Energy Authority, Cekmece Nuclear Research and Training Center Radiobiology Unit Yarımburgaz, Istanbul, Turkey
11. Radiation Protection Center, Vilnius, Lithuania
12. Medical University of Sofia, Sofia, Bulgaria
13. National Center for Radiobiology and Radiation Protection, Sofia, Bulgaria
14. Department of Physiology, Yong Loo Lin School of Medicine: National University of Singapore, Singapore
15. KIRAMS, Seoul, Korea
16. National Research Institute for Radiobiology & Radiohygiene, Budapest, Hungary
17. UNITUS, Viterbo, Italy
18. IMROH, Zagreb, Croatia
19. Fundación para la Investigación del Hospital Universitario LA FE de la Comunidad Valenciana, Valencia, Spain
20. Public Health England, Centre for Radiation Chemical and Environmental Hazards, Chilton, UK
21. Servicio Madrileño de Salud - Hospital General Universitario Gregorio Marañón, Madrid, Spain
22. Canadian Nuclear Laboratories, Radiobiology & Health, Chalk River, Ontario, Canada
23. Federal Office for Radiation Protection (BfS), Oberschleissheim, Germany
24. Serbian Institute of Occupational Health, Radiation Protection Center, Belgrade, Serbia
25. PROCyTOX, Commissariat à l'Energie Atomique et aux Energies Alternatives, Fontenay aux-Roses, France and Université Paris-Saclay, France
26. Institute of Nuclear Chemistry and Technology (INCT), Warsaw, Poland
27. Agenzia Nazionale per le Nuove Tecnologie, L´Energia e lo Sviluppo Economico Sostenibile, Rome, Italy
28. National Center for Scientific Research "Demokritos", NCSR"D", Athens, Greece
29. IRBA, Bretigny sur Orge, France
30. Sri Ramachandra University, Chennai, India
31. Radiobiology Research Unit, Gent University, Gent, Belgium
32. Health Canada, Ottawa, Canada
33. Stockholm University, Institute Molecular Biosciences, Stockholm, Sweden
34. Laboratori Nazionali di Legnaro – INFN, Legnaro, Italy

## Biographical notes:

**Eric Gregoire**, scientist, cytogenetician in biological dosimetry, Institute for Radiological Protection and Nuclear Safety (IRSN), Radiobiology of Accidental Exposure Laboratory (LRAcc), Fontenay aux Roses, France

**Joan-Francesc Barquinero**, PhD, Biologist, University Professor, Department of Animal Biology, Plant Biology and Ecology, Faculty of Biosciencies, Universitat Autònoma de Barcelona (UAB), Bellaterra (Cerdanyola del Vallès), Spain

**Gaetan Gruel**, PhD, Researcher and head of the Laboratory, Institute for Radiological Protection and Nuclear Safety (IRSN), Radiobiology of Accidental Exposure Laboratory (LRAcc), Fontenay aux Roses, France

**Mohamedamine Benadjaoud**, PhD, Biomathematician, Institute for Radiological Protection and Nuclear Safety (IRSN), Radiobiology of Accidental Exposure Laboratory (LRAcc), Fontenay aux Roses, France

**Juan S. Martinez**, PhD, Researcher, Institute for Radiological Protection and Nuclear Safety (IRSN), Radiobiology of Accidental Exposure Laboratory (LRAcc), Fontenay aux Roses, France

**Christina Beinke,** PhD, scientist in the cytogenetics laboratory of the Bundeswehr Institute of Radiobiology, Munich, Germany.

**Adayabalam Balajee,** Head of the Cytogenetic Biodosimetry Laboratory, Radiation Emergency Assistance Center/Training Site, Oak Ridge Institute for Science and Education, Oak Ridge Associated Universities, Oak Ridge, Tennessee, USA.

**Philip Beukes,** Radiation Protection Physicist and Head of Radiation Safety Health Environment and Quality at the National Research Foundation (NRF) iThemba LABS, Cape Town, South Africa.

**William F. Blakely,** senior scientist at his Institute and assistant professor at his University. He is a classically trained radiobiologist and for ~25 years associated with the applied biodosimetry research

and service programs. He is a member of the Scientific Research Department (SRD), Armed Forces Radiology Research Institute (AFRRI) affiliated with the Uniformed Services University of Health Sciences (USUHS), Bethesda, Maryland, United States. He is also the course director for a graduate course in Radiation Biology (PMO-582) at his University, a Council Member of the National Council on Radiation Protection and Measurements (NCRP) serving on program area committee 6 (PAC-6) entitled Radiation Measurements and Dosimetry, and members on International Standard Organization (ISO) Working Groups 18 (Performance criteria for service laboratories performing biological dosimetry by cytogenetics) and 25 (Radiological protection – Radiological monitoring for emergency workers and population following nuclear/radiological accidents – Part 1: General principles).

**Inmaculada Dominguez,** scientist, research in DNA damage and repair, lecturer in Cell Biology, Cell Culture and Radiobiological Group, Cell Biology Department, Faculty of Biology, University of Sevilla, Spain.

**Pham Ngoc Duy, PhD,** Researcher, Biodosimetry Section, Centre of Radiation Technology and Biotechnology, Dalat Nuclear Research Institute, Viet Nam.

**Octávia Monteiro Gil**, PhD, Biology- Genetic, works in the area of radiobiology and biological dosimetry, Instituto Superior Técnico, Centro de Ciências e Tecnologias Nucleares (C2TN/IST/ULisboa), Bobadela, Portugal.

**Inci Güçlü,** scientist, Head of the Radiobiology unit, Turkish Atomic Energy Authority, Cekmece Nuclear Research and Training Center, Istanbul, Turkey

**Kamile Guogyte**, PhD, Chief specialist Radiation Protection Centre Kalvarijų 153 street, Vilnius, Lithuania

**Savina Petrova Hadjidekova**, MD, Assistant Professor, Department of Medical Genetics, Medical University - Sofia, Bulgaria.

**Valeria Hadjidekova**, Director, National Center for Radiobiology and Radiation Protection, Sofia, Bulgaria

**Prakash HANDE,** Associate Professor at the Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore (NUS) and a Fellow at Tembusu College (NUS). biomarkers of radiation exposure, DNA-repair-telomeres-telomerase in ageing and cancer, experimental cancer therapeutics. Dr Hande is one of the pioneers who identified the role of DNA repair factors in telomere regulation in mammalian systems and is an expert in Radiation Biodosimetry. Dr Hande teaches cancer biology and ageing and conducts integrated study module on Biomedicine and Society and Radiation and Society. He holds adjunct professor appointments at the Vellore Institute of Technology, Vellore, India and Mangalore University, Mangalore, India. Dr. Hande is a visiting scientist at the National Institute of Radiological Sciences, Chiba, Japan. He was a consultant at the Division of Human Health, International Atomic Energy Agency, Vienna, Austria in 2015 -2016 while on sabbatical from NUS. Dr Hande is currently an expert member of the workgroup on "Biological mechanisms influencing health effects from low-dose radiation exposure" with United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR).

**Seongjae Jang**, Research scientist in biological dosimetry, National Radiation Emergency Medical Center, Korea Institute of Radiological and Medical Sciences, Seoul, Republic of Korea

**Katalin Lumniczky**, M.D., Ph.D., radiation biologist, head of the Unit of Radiation Medicine, Department of Radiobiology and Radiohygiene, National Public Health Centre, Budapest, Hungary.

**Roberta Meschini**, Research Assistant Professor, Expert in classical and Molecular Cytogenetic and Mutagenesis, Laboratory of Molecular Cytogenetic and Mutagenesis, Department of Ecological and Biological Sciences, University of Tuscia

**Mirta Milić**, scientist, molecular biologist, toxicologist and cytogenetician in biological dosimetry, Institute for Medical Research and Occupational Health (IMROH), Mutagenesis Unit, Zagreb, Croatia

**Alegría Montoro,** PhD in Biology and Master's Degree in Occupational Hazard Prevention, specialising in Hygiene and Security. Supervisor for Radioactive Facilities and head of the biodosimetry laboratory in the Radiation Protection Service (RPS), University-Polytechnic Hospital La Fe,Valencia, Spain

**Jayne Moquet,** Principal Radiation Protection Scientist in the Cytogenetics and Pathology Group, Public Health England - Centre for Chemical Radiation and Environmental Hazards, Oxfordshire, UK.

**Mercedes Moreno Domene**, MSc in radiation biology, Biological dosimetry laboratory. Radiopeathology Centre, Service of Radiation Oncology. Hospital General Universitario Gregorio Marañón (HGUGM-SERMAS), Madrid, Spain

**Farrah N Norton,** Research Scientist-Biologist, lead of the Biodosimetry emergency response capability as well the portfolio lead for the Emergency Response suite of research projects in the Safety and Security program at Canadian Nuclear Laboratories (CNL) in Chalk River, Ontario, Canada.

**Ursula Oestreicher**, PhD, biologist and head of section: "Biological Dosimetry" at the Federal Office for Radiation Protection (BfS), Oberschleissheim, Germany

**Jelena Pajic**, doctor of medical sciences, employed at the Cytogenetic Biodosimetry Laboratory, Serbian Institute of Occupational Health. Main area of research: radiation biology, biodosimetry, genotoxicology.

**Laure Sabatier**, PhD, research director, radiobiologist with molecular cytogenetics expertise, coordinator of biology and health programs and infrastructures at the fundamental research division of the French Alternative Energies and Atomic Energy Commission (CEA)

**Sylwester Sommer**, PhD, radiobiologist, Institute of Nuclear Chemistry and Technology (INCT), Unit: Radiobiology and Biological Dosimetry, Warsaw, Poland

**Antonella Testa,** Antonella Testa, radiobiologist, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Department for Sustainability, Division Health Protection Technologies, Laboratory Health and Environment, Rome, Italy

**Georgia Terzoudi**, physicist and radiobiologist is Director of Research at the Institute of Nuclear and Radiological Sciences & Technology, Energy & Safety, National Centre for Scientific Research

"Demokritos", working in the Health Physics, Radiobiology & Cytogenetics Laboratory in Athens, Greece.

**Marco Valente**, PhD, biologist, cytogenetician, French Armed Forces Biomedical Research Institute (IRBA), Lab: Biological Dosimetry Lab (LDBI), Brétigny-sur-Orge, France.

**Venkatachalam PERUMAL**, Professor in Human Genetics, who is having an extensive background in Radiation Genetics, with explicit training and capability in radiation biodosimetry, bystander response, and genomic instability of high and low dose ionizing radiation, differed in their LET. Sri Ramachandra Institute of Higher Education & Research, Chennai, INDIA.

**Anne Vral**, PhD, full professor and head of the radiobiology research group, principal investigator of the radiobiology group and has 30 years of experience in the field of basic and medically applied radiobiology, radiation protection, biological dosimetry and cancer. The topics related to cancer are dealing with radiosensitivity and DNA repair. An important line of research involves the development and validation of biomarkers of exposure and individual radiosensitivity. Ghent University, Belgium.

**Ruth C. Wilkins**, Research Scientist, Radiobiologist, Health Canada, Ionizing Radiation Health Sciences Division, Ottawa, Canada

**Andrzej Wojcik**, PhD, is professor of radiation biology at the Stockholm University (Sweden) and Jan Kochanowski University in Kielce (Poland). Wojcik focuses on studying cellular effects of radiation, with special focus on factors influencing the radiosensitivity and on combined exposure to radiations of different qualities.

**Demetre Zafiropoulos**, PhD in Biological Dosimetry, italian delegate, of the NEA-OECD Committee on Radiological Protection and Public Health (CRPPH), Radiation Protection Service of Laboratori Nazionali di Legnaro of National Institute of Nuclear Physics.

**Ulrike Kulka**, PhD in biology, head of section national and international cooperation and reporting at the Federal Office for Radiation Protection (BfS) in Oberschleissheim, Germany and chair of RENEB e.V.

## ABSTRACT

### Purpose

In case of a mass-casualty radiological event, there would be a need for networking to overcome surge limitations and to quickly obtain homogeneous results (reported aberration frequencies or estimated doses) among biodosimetry laboratories. These results must be consistent within such network. Inter-laboratory comparisons (ILCs) are widely accepted to achieve this homogeneity. At the European level, a great effort has been made to harmonize biological dosimetry laboratories, notably during the MULTIBIODOSE and RENEB projects. In order to continue the harmonization efforts, the RENEB consortium launched this intercomparison which is larger than the RENEB network, as it involves 38 laboratories from 21 countries. In this ILC all steps of the process were monitored, from blood shipment to dose estimation. This exercise also aimed to evaluate the statistical tools used to compare laboratory performance.

### Materials and Methods

Blood samples were irradiated at three different doses, 1.8, 0.4 and 0 Gy (samples A, C and B) with 4-MV X-rays at 0.5 Gy $min^{-1}$, and sent to the participant laboratories. Each laboratory was requested to blindly analyze 500 cells per sample and to report the observed frequency of dicentric chromosomes per metaphase and the corresponding estimated dose.

### Results

This ILC demonstrates that blood samples can be successfully distributed among laboratories worldwide to perform biological dosimetry in case of a mass casualty event.

Having achieved a substantial harmonization in multiple areas among the RENEB laboratories issues were identified with the available statistical tools, which are not capable to advantageously exploit the richness of results of a large ILCs. Even though Z- and U-tests are accepted methods for biodosimetry

ILCs, setting the number of analyzed metaphases to 500 and establishing a tests' common threshold for all studied doses is inappropriate for evaluating laboratory performance.

Another problem highlighted by this ILC is the issue of the dose-effect curve diversity. It clearly appears that, despite the initial advantage of including the scoring specificities of each laboratory, the lack of defined criteria for assessing the robustness of each laboratory's curve is a disadvantage for the "one curve per laboratory" model.

Conclusions

Based on our study, it seems relevant to develop tools better adapted to the collection and processing of results produced by the participant laboratories. We are confident that, after an initial harmonization phase reached by the RENEB laboratories, a new step towards a better optimization of the laboratory networks in biological dosimetry and associated ILC is on the way.

## 1- INTRODUCTION

In case of a large-scale radiation emergency, the dose estimation of the victims should be assessed as fast and accurate as possible. Firstly, a triage should be performed by qualified medical staff according to clinical signs. Subsequently, a categorization of the exposed people by dosimetry must be carried out according to their degree of suspected overexposure. Biological dosimetry rapid assessment complements the clinical triage by categorizing potentially exposed victims in different ranges of exposure (Vaurijoux et al. 2015; Ainsbury et al. 2014) and is a key element when physical dosimetry is not available (Christie et al. 2010; Romm et al. 2014a). Because triage based on manual dicentric chromosome analysis (DCA) is done with a low number of analyzed cells (usually 50), it makes it highly imprecise as it has large confidence intervals. For this reason, the categorization should be defined by dose ranges and not in terms of dose alone. Furthermore, after initial triage, dose assessment is needed in order to confirm the categorization and to give a more precise individual dose estimation (Romm et al. 2014a). As precise dose assessment requires the analysis of a large number of cells, usually from 500 to 1000 by DCA, the time needed for a correct dose assessment is larger than that needed for triage. As an example, using manual scoring, one operator would perform triage in 1 hour per sample, but dose estimation would take approximately two days.

In general, biodosimetry laboratories can manage only a limited number of victims at one time. Thus, in the case of a mass-casualty radiation incident, where the management of several hundreds of victims would need to be performed, there is a prerequisite for national and/or international networking. However, networking must be based on the ability to provide homogeneous results (Voisin 2015; Kulka et al. 2015, 2017). This means that for any single case, the reported chromosomal aberration frequencies or estimated doses should be consistent and comparable among the laboratories responding to the emergency. Harmonization needs standardized procedures; this is an essential point for the successful coordination of different laboratories (Beinke et al. 2013; ISO 19238; Wilkins et al. 2008; Christie et al. 2010; Beinke et al. 2011). The strategy of establishing a cooperative network among laboratories requires that each laboratory follows internationally accepted methods for analysis (IAEA 2011, ISO 19238 2014) and regular inter-laboratory comparisons (ILCs) to test performance analysis (Wilkins et

al. 2008, Di Giorgio et al. 2011). Nowadays it is widely accepted that networking should include regular international ILC exercises simulating different scenarios, as this would guarantee a more rapid response and a higher reliability of dose estimates (Wojcik et al. 2010).

During the last decade several ILCs have been performed. Some of them were focused on the triage (Wilkins et al. 2011; Lloyd et al. 2000; Ainsbury et al. 2009; Garcia et al. 2013; Romm et al. 2011, 2014a, b, Oestreicher et al 2017) while others mainly on dose-assessment (Yoshida et al. 2007; Pan et al. 2019; Bakkiam et al. 2015; Liu et al. 2016, Roy et al. 2004). In a large-scale ILC involving 7 countries from the Latin American Biological Dosimetry Network (LBDNet) and 6 laboratories from the European Union, a good agreement among participants was shown in terms of the reported dicentric chromosome yields and assessed doses. In this ILC the results after the analysis of 50, 100 or 500 cells from shared stained slides were evaluated by using robust methods described in different ISO standards (Di Giorgio 2011). Another effort in validating international networking using the DCA in the case of a potential mass-casualty event was done by Wilkins et al. (2008). Several ILCs based on triage have shown that more than 90% of the participant laboratories correctly categorize the tested samples (Miller et al. 2007; Di Giorgio et al. 2011; Beinke et al. 2011, 2013; Bhavani et al. 2014; Yoshida et al. 2007; Roy et al. 2004).

At European level, different projects have been founded in order to improve standardization and harmonization for the different biomarkers of dose. MULTIBIODOSE helped in defining what would be the best assay to use depending on different exposure scenarios (Jaworska et al. 2015, Ainsbury et al. 2014). In addition, a NATO project studied the possibility of reducing the number of analyzed cells from 50 to 20 for triage purposes (Beinke et al. 2013). Recently, several RENEB (Realising the European Network of Biodosimetry) project training sessions and ILCs have allowed the main cytogenetic assays to be homogenized and standardized among participants. Therefore, RENEB has helped in creating an efficient European network of biodosimetry laboratories (Kulka et al. 2012). The harmonization and the quality of the results for triage mode obtained among the RENEB members let us claim that at the present day RENEB is able to categorize a large number of victims in mass-casualty radiological events (Kulka et al. 2017; Gregoire et al. 2017; Oestreicher et al. 2017).

As indicated above, an individual dose assessment should be performed after triage (Romm et al. 2014a, Wojcik et al. 2010). The present ILC concerned different biomarkers (Micro-Nucleus (MN), Premature Chromosome Condensation (PCC), gamma H2AX and Gene Expression) and its main goal is to delve deeper and check the entire process needed for proper networking, from blood sample shipment to dose estimation. The present work has been focused on the DCA. In addition to the 20 RENEB laboratories, another 19 laboratories were invited to participate. Finally, as this ILC evaluates the entire process for dose-assessment by biodosimetry, we will take advantage of the large data set to critically review the statistical tools used to evaluate laboratory performance.

## 2- MATERIALS and METHODS

### a. Irradiation and Shipment

A 420 ml blood sample from a female donor (Etablissement Français du Sang (EFS), France; Agreement CPSL UNT N°13/EFS/123) was irradiated at 37°C in a water bath with 4-MV X-rays delivered by a Linear medical accelerator (Elekta Synergy, IRSN, Fontenay aux Roses, France) at 0.5 $Gy \cdot min^{-1}$, dose in water. The irradiation field was 30 x 30 cm and the distance between the source and the sample was of 1.07 m. Radiation field mapping and dosimetry was confirmed using cylindrical ionization chamber (0.125cc n° 4920) calibrated in dose to water. The blood sample was placed in 3 tubes corresponding to the different dose points, a high dose of 1.8 Gy, a low dose of 0.4 Gy, and a sham-irradiated sample. After irradiation, samples were maintained 2 h at 37°C and then the blood was aliquoted into 2 mL tubes. Blood samples were then coded as follows: the high dose as A, the low dose as C and the sham irradiated as B. Then, samples were sent to the 39 participant laboratories from 19 countries who were informed by e-mail of the shipment of three samples. The e-mail informed that there were three blind samples, that corresponded to ~~a~~ high-, low- and sham- irradiated samples. In the same e-mail the RENEB standard scoring sheet for dicentrics, or dicentrics plus centric ring, analysis was attached.

Shipment was performed using commercial express delivery services as UN 3373 Biological Substance Category B, as described in detail in the manual of the International Atomic Energy Agency (IAEA

2011). Each package of blood samples included a temperature logger and a dosimeter to monitor the temperature and any dose received during transportation. A second blood sample (male donor, Etablissement Français du Sang (EFS), France; Agreement CPSL UNT N°13/EFS/123) was processed as above and sent to 2 laboratories for whom there were shipment issues (see section 3).

In this study, the laboratories classified as RENEB laboratories (L1 to L20) correspond to those belonging to the RENEB project that took part in the last RENEB ILC in 2015 (Oestreicher et al. 2017). The other participants are classified as non-RENEB group (L21 to L38).

**b. Cell culture and dicentric chromosome assay**

Thirty of the participant laboratories were requested to set up lymphocyte cultures. Blood samples were transmitted to three other participants by an intermediary laboratory in Bulgaria or South Korea. Thus 33 laboratories received blood samples. In all cases, cultures were processed using each laboratory's standard protocol following the recommendations of the IAEA (2011) and the ISO standard 19238 (2014). Finally, a contact laboratory from Canada set up the lymphocyte cultures and sent stained slides to its network of 6 laboratories. In all cases, the analyses were performed according to a RENEB standard scoring sheet for the dicentric chromosome assay that was provided to the 39 participants. For each sample, manual scoring of dicentric chromosomes (or dicentric chromosomes plus centric rings) in 500 cells by two different scorers if possible and using at least two slides (250 cells in each) was requested. In addition to dicentric frequency per metaphase and dose assessment (Gy) for each sample, participants were asked to report the Colcemid treatment used and the coefficients and associated errors of their calibration curve. All participants sent the results directly (30 laboratories) or indirectly (through their reference laboratory in Bulgaria, Canada and South Korea) to the organizing laboratory at the IRSN.

**c. Dose assessment**

For dose assessment, laboratories converted the frequency of aberrations observed per metaphase into absorbed dose using their own calibration curves based on dicentric chromosomes or dicentric chromosomes plus centric rings scoring. Some of the laboratories without their own calibration curve decided to use the calibration curve data available in the last technical IAEA report on biodosimetry

(IAEA, 2011). The calculations needed to convert the observed dicentric chromosomes (or dicentric chromosomes plus centric rings) into estimated doses were made by means of various software programs: CABAS V2.0 (Deperas et al 2007), different versions of Dose Estimate (Ainsbury and Lloyd 2010), or in-house developed software based on Microsoft Excel (L6, L15, L19 and L34). All laboratories estimated the doses in Gray (Gy) and included the 95% confidence intervals as requested.

**d. Statistics**

To assess the performance of each laboratory and the reproducibility of the exercise, the statistical analysis followed the (ISO 5725 1998) recommendations which provide detailed guidance of general statistical methods to use in proficiency testing schemes. These methods were successfully applied for biological dosimetry in Di Giorgio et al. (2011). In brief, the robust estimations of the mean and standard deviations of frequencies or doses were performed using the Algorithm A (algA function of the R software "metrology" package) (ISO 13528, 2015). This algorithm yields robust location and scale estimates by the "winsorisation" of the original data (the extreme values, instead of being deleted, are shifted towards the bulk of the data using adequate upper and lower thresholds obtained by an iterated scale). The "Breakdown points" for these estimators (proportion of outliers without an adverse impact on the estimates) are approximately 30, which constitutes an adequate resistance to outlying values. A robust estimation of the coefficient of variation can then be obtained as the ratio of the robust standard deviation to the robust mean.

Once the mean and standard deviation robustly estimated, the performance analysis was conducted using the Z- and U-tests. The Z-test measures the deviation of each laboratory's reported frequency or estimated dose from the robust mean of the reported frequencies or the delivered dose, both considered as reference values. The Z-test also takes into account a robust standard deviation from the reported frequencies or doses, and a standard uncertainty of the reference value. Laboratory performance using the Z-test categorizes reported values into "satisfactory" when the $|Z|$ value is $\leq 2$, "questionable" for a $|Z|$ value between 2 and 3, and "unsatisfactory" when the $|Z|$ value is $\geq 3$. Z-tests do not consider the uncertainty of each participating laboratory. On the other hand, the U-test considers the mean value and its confidence interval. With the U-test, the results of each laboratory are interpreted considering the

upper critical value of Student's t distribution, usually with a 0.05 probability of exceeding the critical value, and with N-1 degrees of freedom (where *N* is the number of laboratories). For both tests, Z and U, and to prevent against the multiple testing issues in the statistical inference, the Benjamini-Hochberg (BH) (Benjamini and Hochberg 1995) adjustment was performed for controlling the false discovery rate (FDR). This FDR-based control has been widely used in cases where a large number of hypotheses are simultaneously tested and has been shown to be less conservative than the Bonferroni adjustment (Shaffer 1995).

## 3- RESULTS

### a. Shipment

A total of 39 laboratories were involved in this ILC. From the initial shipping of blood samples by the organizing laboratory (IRSN), 11 laboratories belonging to the European Union (EU) received them within a period of 24h; and 22 laboratories received them within a period of 48 h regardless of location (EU or non-EU). 4 laboratories received the blood samples after a period larger than 48 h (from 48.5 h to 68 h), and 2 laboratories did not receive the samples due to an issue with customs authorities. A new shipment for the latter 2 was made by the IRSN laboratory and it was received within 48 hours. Each package included a dosimeter and none of the recorded doses was above 0.1 mSv. Among the laboratories that received blood samples, 32 out of 33 were able to set up lymphocyte cultures and to successfully obtain chromosome spreads. Considering all participants, including the 6 labs that received coded slides with chromosome spreads, a total of 38 laboratories were able to report their results.

### b. Reported Frequencies

Table 1 shows the total number of dicentric chromosomes or dicentric chromosomes plus centric rings found by each laboratory for the three evaluated samples (A, B and C). From the total 114 reported values (38 labs x 3 samples), 99 of them correspond to the analysis of around 500 cells and the other 15 values correspond to 300 analyzed cells or less (Table 1). Two laboratories submitted 2 dose-effect curves each for manual scoring, as follows: one sent curves based on different kinds of staining (Giemsa (L2) or FISH coupled with pan-telomeric and pan-centromeric probes (L2b)), and one sent curves based on chromosomal aberration scoring (dicentrics (L31) and dicentrics plus centric rings (L31b)). All the sent results have been included in the analysis to show what can happen in a real case. Indeed it is possible, whatever the cause, that a laboratory obtains a poor mitotic index, even when it is a very experienced laboratory.

TABLE1NEARHERE

The figure 1 shows only the dicentric frequency per metaphase in order to effectively compare the same frequencies to each other. For sample A (1.8 Gy), the dicentric frequencies sent by each participant

laboratories are shown in Figure 1A. The observed frequencies of dicentric chromosomes per cell ranged from 0.10 to 0.34. The robust estimate of the mean (± robust standard deviation) was 0.22 ± 0.058, and the coefficient of variation (CV) was of 27%. When only RENEB laboratories were considered CV was 20%. Figure 1 also shows the results of the Z and U tests.

FIGURE1NEARHERE

Z- and U-tests were only done using the frequency of reported dicentric chromosomes per metaphase (Figure 1). For both tests, the BH adjustment was applied. The Z-score obtained for 97% of the labs were satisfactory. Different results were obtained using the U test, where the results of 10 labs (26% of labs) were unsatisfactory after BH adjustment. Evaluating separately RENEB (from L1 to L20) and non-RENEB laboratories (L21 to L38), we can notice that 85% of RENEB laboratories had satisfactory U-test values and among non-RENEB laboratories, only 61% showed satisfactory U-scores.

For sample B (0 Gy), dicentric chromosome frequencies sent by each participant laboratories are shown in Figure 1B. The observed frequencies of dicentric chromosomes per cell ranged from 0.0 to 0.01 and the robust estimate of the mean (± robust standard deviation) was 0.0014 ± 0.0017 and the CV was 128%. The results of the Z and U tests are also shown in Figure 1B. Concerning the Z-test, the results of only three participants (L14, L20 and L29) were considered unsatisfactory. 90% of the RENEB and 94% of the non-RENEB participants had a satisfactory Z-score. The U- test considered that all the values given by the laboratories are satisfactory.

The dicentric frequencies for sample C (0.4 Gy) sent by each participant laboratories are shown in Figure 1C. Observed mean frequencies of dicentric chromosome per cell ranged from 0.0 to 0.08 and the robust estimate of the mean (± robust standard deviation) was 0.025 ± 0.011. The CV was 44%.

As shown in Figure 1C, only one Z value was considered as unsatisfactory (L7). Therefore, 97% of the participants present satisfactory Z-scores. When the U-test was applied, 2 frequencies gave unsatisfactory results (L36 and L37). With the U-test, L7 was not unsatisfactory anymore as the

uncertainty associated with its frequency per metaphase is quite large due to the analysis of only 12 cells. So, 95% of the participants had a satisfactory U-score.

### c. Estimated Doses

TABLE2NEARHERE

The second step of this intercomparison was to estimate the three delivered doses. The ILC requested each laboratory to calculate the estimated doses and their associated confidence intervals, using their own dose-effect curve and applying the statistical method established in their laboratory. In addition, the RENEB scoring sheet requested each laboratory to indicate the coefficients and standard errors of the calibration curve used (Table 2). Twenty-nine laboratories sent the coefficients of a single dose-effect curve, generally constructed using gamma- or X-rays (Table 3). Four laboratories submitted 2 dose-effect curves each, as follows: 2 participants sent curves based on different irradiation sources (gamma- (L4 and L5) and x-rays (L4b and L5b)), and the two others (L2/L2b and L31/L31b) were mentioned earlier in the "reported frequencies" section. Finally, five laboratories did not have any calibration curve but two of them chose to use the calibration curve data available in the last technical IAEA report on biodosimetry (IAEA, 2011), as it is also an acceptable method. For the other three, as it was their first time of participation to an intercomparison, they were not aware of the possibility to use an established dose-effect curve. Figure 2 shows only the values from the participants that reported an estimated dose.

TABLE3NEARHERE

FIGURE2NEARHERE

Using the values sent by each laboratory, 65% of the reported dose estimation participants include the high dose (sample A) in the 95% confidence interval of their dose estimates (93% and 89% for sham irradiation (sample B) and low dose (sample C), respectively). As mentioned above, each laboratory calculated the absorbed doses using the program routinely used in their laboratory, and there was a great heterogeneity in the calculation of the 95% confidence interval. In fact, 16 laboratories used the CABAS

software that only considers Poisson's error on the observed yield. 13 participants used Dose Estimate software, that can consider both, the error of the curve and the error of the observed yield of dicentric chromosomes applying the delta method (IAEA manual 405, 2001). Among the 13 participants which used the Dose Estimate software, 11 considered the delta method, and 2 only considered the error of the observed yield. As well 6 laboratories that used their own software applied a Merkle approach to consider both errors (Merkle, 1983). Finally, 3 laboratories gave no results on dose estimation. Among the 38 laboratories that sent results, 2 laboratories sent miscalculated doses due to typo errors. To avoid the impact of this heterogeneity in Z- and U-test analysis, all the dose estimations were recalculated using a single method. The method used was Merkle's approach that was proposed in the last IAEA manual (IAEA, 2011). However, because covariances of the fitted coefficients of curves were not previously requested, the 95% confidence intervals were calculated considering only the standard errors on curve coefficients. These results are reported in Figures 3, which show 39 results each because some laboratories provided 2 dose effect curves leading each to dose estimations.

FIGURE3NEARHERE

For the high dose (sample A), the estimated doses ranged from 1.31 to 2.51 Gy, and 90% of the participants included the delivered dose in their 95% confidence interval (Figure 3A). After applying the BH adjustment all laboratories showed satisfactory Z-scores. Using the U-test, 90% of the laboratories showed satisfactory results, 96% for RENEB and 81% for NON-RENEB participants. The CV was of 15%.

For the sham-irradiated sample (B), the estimated doses ranged from 0.0 to 0.19 Gy, and in all cases the 95% confidence intervals included the 0 Gy dose, except L35 (Figure 3B). For sample B, results cannot be analyzed using the Z-test because of the algorithm A convergence failure of the robust standard deviation estimation (abundance of null values). The U-test showed that 97% of the results were satisfactory. All RENEB and 94% of non-RENEB participants had satisfactory U-scores.

Results for the low dose (sample C) can be seen in Figure 3C. Estimated doses ranged from 0.24 to 1.20 Gy, and the Z-test shows unsatisfactory result for only one participant. Therefore, 97% of the laboratories had satisfactory scores. All U-scores were satisfactory. The CV was 29%.

Importantly, we noticed a substantial heterogeneity in the calibration curves from the participants as reported in Table 3 and Figure 4. Concerning the gamma-rays calibration curves, the lowest dose rate was 0.04 $Gy \cdot min^{-1}$ (L3), and the highest 1.16 $Gy \cdot min^{-1}$ (L34). For calibration curves using X-rays, dose rates ranged between 0.35 $Gy \cdot min^{-1}$ irradiating with X-rays of 243 kVp (L4b) and 2.5 $Gy \cdot min^{-1}$ irradiating with X-rays of 6 MeV (L19). The IAEA technical report (IAEA, 2011) recommends that to produce a dose-effect curve applicable to an acute accidental exposure the dose rate should be chosen such that all doses are given in less than 15 min. Considering this recommendation and taking into account that usually the highest dose used in a calibration curve is 4 or 5 Gy, a dose rate of about 0.34 $Gy \cdot min^{-1}$ will allow to irradiate the highest dose in less than 15 minutes. Therefore 7 laboratories (L3, L7, L16, L22, L33 and L34) used a dose rate that is under the IAEA recommendations.

An alternative to reporting satisfactory result rates is to rank the results of each laboratory belonging to the same network for a given sample based on their Z- and U-scores. Table 4 shows the differences in ranking of the laboratories between the 2 tests.

TABLE4NEARHERE

In fact, differences in scoring criteria should be balanced by the use of individual curves, which logically includes the specific scoring criteria of each laboratory. This effect is not clearly observed in the present study. Table 5 shows the differences of laboratory ranking from the frequency to the dose estimation by Z-score. For example, at group level, the mean Z-score calculated for the RENEB network or the non-RENEB group does not change as much between frequency and estimated dose. For frequency and dose estimation, RENEB and non-RENEB laboratories were ~~ranking~~ ranked based on their Z-scores, from the lower to the higher values. The mean of the rank obtained for RENEB and non-RENEB laboratories are 17.4 and 21.8 respectively. This is quite the same for dose estimation, the mean of the laboratory

rank based on Z-scores is 18.7 for RENEB, and 21.8 for non-RENEB laboratories. It should be noted that the mean dose for all laboratories is not far from the delivered dose (1.74 Gy vs 1.80 Gy).

TABLE5NEARHERE

The curves reported for the present ILC show great variability in their calibration curve coefficients (Table 2) and highlight the existing diversity among laboratories. A more visual representation of these differences can be seen in figure 4. As an example, a frequency of 0.5 dicentric chromosome per metaphase gives a dose of 1.80 Gy for L2b and a dose of 3.90 Gy for L9.

FIGURE4NEARHERE

## 4- DISCUSSION

In a mass-casualty radiological event, networks of biological dosimetry laboratories can decide to share different types of samples such as whole-blood, fixed cells, slides or metaphase images. Multiple Inter-Laboratory Comparisons (ILCs) have already tested different possibilities: blood (Roy et al. 2004; Oestreicher et al. 2017, Bakkiam et al. 2015, Romm et al. 2011, Pan 2019, and Wilkins et al. 2008); fixed cells (Roy et al. 2004); slides (Liu et al. 2016; Miller et al. 2007); or metaphase images (Garcia et al. 2013, Romm et al. 2014a and 2016). The present ILC has chosen to send whole blood to test all the steps of a biological dosimetry study (i.e. blood culture, slide preparation and staining, dicentric analysis and dose estimation).

Evaluating the shipment, 87% of the participant laboratories received the blood samples within 48h, including those outside Europe (Canada, USA, South Africa, South Korea, India and Vietnam). In addition, 97% of the laboratories were able to obtain chromosome spreads, even those that received the samples after 48h. In fact, only 6 out of 38 laboratories did not reach the 500 metaphases needed. The only laboratory that did not obtain any chromosome spreads received the blood sample in 48h. Therefore, no link could be established between sample-travel time and culture growth, and some delay in the shipment did not prevent lymphocyte growth in this study. In future ILCs it would be of interest to report the mitotic index in order to evaluate lymphocyte activation. Moreover, the impact of the shipment itself has been tested in other exercises. Particularly, in the ShipEx exercise between the Latin-American network (LDBNet) and several laboratories around the world. In this case, blood samples were also properly received and lymphocytes were able to satisfactorily grow for most of the participants (Garcia et al. 2013). The blood shipment has also been tested in other European ILCs (Beinke et al. 2013), where the same observation was made for the longest shipment times, including 96 h but this was not optimal (Oestreicher et al. 2017).

In our study, the dose received during the transport of the samples (mainly cosmic radiation and X-ray safety checks) amounted to a maximum of 0.1 mSv. In addition, only long-distance shipments were exposed to measured doses between 0.05 and 0.1 mSv. As comparison, similar exposures were reported in the ShipEx-1 exercise (Garcia et al. 2013). Thus, these dose levels can be considered negligible

compared to the sensitivity limits of the biological dosimetry method used. However, it seems prudent to systematically include a dosimeter in the blood sample shipment, in case there are abnormal exposure levels during transit safety checks. The results presented here and those previously reported (Oestreicher et al. 2017, Wilkins et al. 2008) show that blood samples could be shared among laboratories around the world in the event of a major radiological accident in order to perform biological dosimetry based on chromosome aberrations.

**Interpretation of ILC Results**

Periodic ILCs allow the evaluation of the performance of laboratories that belong to a network. They help to standardize practices and contribute to the improvement of the quality and robustness of the results from such a network. One important aim of ILCs is to identify problems encountered by the participants and define actions for improvement, such as harmonization, training and dose estimation exercises. In biological dosimetry, the results are mainly based on the estimation of the chromosome aberration frequency per metaphase. This value is subsequently converted into an estimated absorbed dose using a pre-established dose-response curve specific to each laboratory. For biodosimetry laboratories, the main goal of an ILC is to compare the results for these two values, frequency and estimated dose, among the participant laboratories.

The objectivity of these comparisons is generally achieved through the Z- and U- score tests (Di Giorgio et al. 2011). These two quantities evaluate, under different normalizations, the difference between the value reported by each laboratory and a reference value considered as correct (i.e. the robust mean for frequencies, or the delivered dose for dose estimation). In fact, the Z-Score is computed under a common normalization based on the robust standard deviation while the U-score is computed using a laboratory specific normalization based on the uncertainty measurement of each participant (which is highly associated to the number of cells scored, but also to the level of exposure). Thus, these two tests give complementary elements to interpret ILC results. In the present study, the U- and Z-scores were adjusted using the Benjamini & Hochberg (1995) correction in order to take into account the large number of calculated scores (at least one for each participating laboratory) and the associated increase of false

positive risk. Finally, by defining thresholds, one could distinguish acceptable, questionable and unsatisfactory results. The comparison of frequencies aims to provide an overview of the state of harmonization between the participating laboratories concerning chromosome aberration recognition. In other words, this analysis allows the evaluation of the homogeneity among participating laboratories concerning aberration detection and scoring criteria. From an overall perspective, with the Z-score, the percentage of satisfactory results decreases with the level of exposure: 100%, 97% and 92% for samples A (1.8 Gy), C (0.4 Gy), and B (non-exposed) respectively. Contrary to what the percentages of satisfactory results might suggest, it cannot be concluded solely on the basis of the Z-score that laboratory harmonization is worse at low doses than at high doses. The reason is that these percentages are simply not directly comparable. In fact, the standard ISO 13528 (2015) justifies the use of the limits "2" and "3" for the Z-score by the fact that "measurements that are carried out correctly are assumed to generate results that can be described by a normal distribution". Therefore, it is easy to see that the validity of the Z-scores limits (2 and 3) is intrinsically related to the large-sample asymptotic normal approximation of a Poisson distribution, which is usually used to describe the distribution of dicentric chromosomes in a uniform irradiation context. The Berry-Essen Theorem (Berry 1941, Essen 1942) provides an easy way to quantify this convergence rate which, in the case of a Poisson distribution, states that a bound on the maximal error of the normal approximation is inversely proportional to the square root of the product of the number of metaphases times the dicentric rate.

According to the sample A and the sample C aberration rates per cell (approximately 0.2 and 0.02 respectively), this implies that 5000 analyzed metaphases are needed for the low dose (sample C) to achieve the normal approximation precision after analyzing 500 metaphases of the high dose (sample A). In other words, by fixing the number of analyzed metaphases to 500 for all investigated doses, the corresponding Z-score distributions are significantly different in terms of normal approximation, making it inappropriate to have common satisfactory/unsatisfactory thresholds (here 2 and 3).

The same conclusion can be made for the U-score, even though it gives opposite results. With the U-test, the percentage of satisfactory results decreases when the sample dose increases: 100% for non-irradiated sample (dose B), 95% for 0.4 Gy (dose C) and 74% for 1.8 Gy (dose A). Once again, and for

the same reasons explained above for the Z-scores, it would be erroneous to conclude that laboratory harmonization is worsening as the dose to be estimated increases.

An alternative to reporting satisfactory result rates is to rank the results of each laboratory belonging to the same network for a given sample based on their Z- and U-scores. As explained above, the methodology underlying the Z- and the U-score is not the same and the analysis of the ranking obtained with each one should be interpreted in light of these differences. The Z-score ranks the laboratories based on the distance between the value reported by each of them and the reference frequency (i.e. the robust mean of all reported frequencies). Basically, the farther you are from this average value representative of the group, the lower you are ranked. A disadvantage of this method is that two laboratories that report the same mean frequency will have the same score, even if one of them has a larger uncertainty for the measurement. This can be illustrated by comparing the Z-score of L6 and L7 for sample A. The two laboratories obtained a Z score very similar (0.34 and 0.28) as their reported frequencies for sample A are similar (Table 4). However, the frequency of L7 has a much higher uncertainty due to the low number of scored metaphases, and it can be considered less reliable than the result of L6, which is not reflected in the Z-score ranking. The U-score makes it possible to account for this difference between the 2 laboratories, but not in the direction that one would expect. In fact, the U score for L7 (0.12), is lower than for L6 (0.86). Therefore, when performing the U-test for two laboratories with similar frequencies, one of them can be better ranked because of its large uncertainty. Because the number of dicentric chromosomes that can be detected will depend on the delivered dose and on the number of cells analyzed, these two tests should be used carefully when ILC frequencies of detected aberrations are considered. It seems more reasonable to use these tests to evaluate the level of harmonization between laboratories, or networks of laboratories, rather than to evaluate each laboratory's performance. The present RENEB ILC involves laboratories belonging to different groups (RENEB network and non-RENEB participants) that have independently harmonized dicentric chromosome scoring. In the present ILC, RENEB laboratories constitute ~~majority~~ half of the participants and most of them have already participated to several ILCs (Oestreicher et al. 2017, Jaworska et al. 2015, Romm et al. 2014a, Ainsbury et al. 2014). This has a strong effect in the robust mean and robust standard deviation considered as reference values.

The global analysis of frequencies by Z-score and U-score as presented in Table 4 can lead to misinterpretations. As mentioned above, the real interest of the frequency analysis is to evaluate the level of harmonization in the recognition of dicentric chromosomes. The present study involving 38 laboratories around the world that do not belong to the same network, or do not even belong to a network, necessarily highlights an expected heterogeneity in the results, which is not very surprising or informative. It might be of interest to focus on a sub-group of laboratories that have worked to harmonize themselves, and to assess the gain associated with this harmonization process.

TABLE6NEARHERE

Table 6 presents the results for sample A and for the Z-score analysis performed only on the 20 laboratories belonging to the RENEB network. Within this group, the robust coefficient of variation is 20.1%, with 3 labs (15% of all RENEB labs) showing questionable results (L11, L18 and L3). If these 3 laboratories are excluded, the coefficient of variation calculated from the frequencies obtained by the remaining 85% of the laboratories is around 15%. These values can then be compared to the expected value for the coefficient of variation which can be obtained by simulating 20 or 17 chromosome aberration frequency estimates following a Poisson distribution with a parameter (lambda) equal to the robust means observed on the RENEB subgroup, and taking into account the respective numbers of metaphases scored by each laboratory. Then, the median value of these “theoretical” coefficients of variation is 13.5% with 95% confidence interval of [7.7% - 27.4%]. This means that, due to the stochastic nature of the measures, 20 laboratories involved in a "fully harmonized" ILC situation is expected to obtain, in median, a coefficient of variation of 13.5%. Thus, the dispersion of the values obtained for the RENEB network, 20% or 15% is included within the 95% confidence interval of the “theoretical” coefficient of variation and close to the median “theoretical” value of 13.5%. In comparison, the robust coefficient of variation obtained for all 38 laboratories is 26.7%, and if only the non-RENEB laboratories are considered, the dispersion reaches a value of 36.5%. This shows that an intercomparison analysis based on chromosomal aberration frequencies only makes sense among laboratories that are involved in a common effort of harmonization. This should not be interpreted as a

better proficiency of a specific network in detecting dicentric chromosomes with respect to another, but as a reflection of different ways of harmonization.

In conclusion, although Z- and U- tests are accepted methods to assess laboratory performance in metrology (ISO 13528 2015), they are not ideal for ILCs. To this day, no commonly used tool proves to be fully adapted and relevant to the needs of ILCs that are based on the frequency of radiation-induced dicentric chromosomes per metaphase. To mitigate this deficiency, it seems appropriate for the reliability of future ILCs to focus on radiation doses that are able to generate enough dicentric chromosomes for 500 analyzed cells. This would limit the impact of Poisson uncertainties on the ILC results. In addition, it seems essential to only include in the intercomparison analysis those laboratories that have analyzed the requested number of metaphases, and to exclude those that have not, thus allowing a comparison with an equivalent Poisson uncertainty. Otherwise, a comparison of results from all participants appears hazardous. Additionally, one should consider that ILCs may include laboratories from different networks that could have their own harmonized way of scoring dicentric chromosomes. This could lead to questionable or unsatisfactory results because of different scoring criteria, and it should not be interpreted as bad performance, but as a lack of harmonization among all the participating laboratories.

While ILCs based on the frequency evaluate the level of harmonization of scoring criteria, intercomparisons based on estimated doses involve additional elements to be taken into account when analyzing the results. One is the dose-effect curve required to estimate a dose from the observed frequency, of which most of the participant laboratories have their own. It is widely accepted that in ILCs, better results are generally obtained with estimated doses than with observed frequencies (Di Giorgio et al. 2011). In fact, differences in scoring criteria should be balanced by the use of individual curves, which logically includes the specific scoring criteria of each laboratory. This effect is not clearly observed in the present study. For example, at group level, the mean Z-score calculated for the RENEB network or the non-RENEB group does not change as much between frequency and estimated dose. For frequency, the mean rank based on the Z-scores obtained for RENEB and non-RENEB laboratories are 17.4 and 21.8 respectively. This is quite the same for dose estimation, the mean rank based on the Z-

scores are 18.7 for RENEB, and 21.8 for non-RENEB. This is confirmed at the laboratory level, as drastic changes in Z-score ranking (gain or loss of more than 15 places) between frequency and dose estimation are limited to a minority of laboratories (Table 5). This is the case for L18, which systematically improves its Z-score by more than 1.5 points between frequency and estimated dose (a progress of 25 ranking places). Inversely, L9 gained more than 2 Z-score points when its dose-effect curve was used to convert its chromosome aberration frequency to an estimated dose (a 35-row drop in the overall ranking).

It is interesting to note that, although these changes in results between frequency and dose are small in magnitude for most laboratories, globally, they are quite unfavorable. In fact, an increase of the Z-score is observed for the majority of participants (58% for the high dose, sample A and 55% for the low dose, sample C) when estimated doses are considered. Indeed, for the high dose, a mean loss of 2 ranks per lab were observed between the ranking obtained for frequency and the one obtained for dose estimation. This could indicate that the dose-effect curves include biases that prevent them from positively compensating for differences in scoring criteria. The curves reported for the present ILC show great variability in their calibration curve coefficients (table 2) and highlight the existing diversity among laboratories. A more visual representation of these differences can be seen in figure 4.

The above-mentioned differences have multiple origins such as the number of dose points used to calibrate the curve, the number of metaphases analyzed at each dose point, the dose-rate and the radiation source (X- or gamma-rays). Another source of uncertainty is the way that the delivered doses were calculated (Trompier et al, 2017). Briefly, depending on the radiation source, X-or gamma-rays, and their energy, the calculation of the delivered dose to the samples can be based on air Kerma or dose to water. Depending on the overall energy of the source, this could lead to different absorbed dose values for the same irradiation. Consequently, this can impact the result of the dose estimation in an ILC if the doses of a given dose-effect curve are not calibrated the same way than the dose delivered to the analyzed sample. It is important to mention that usually in biodosimetry laboratories, all these details are not very well traced. Furthermore, there are no minimum criteria for defining whether or not a dose-response curve is acceptable for use in a given intercomparison. Currently, and in most of ILCs, calibration curves from all participants are used, regardless of the way they are built. In fact, 7 laboratories reported dose

effect curves that were built using a dose rate too low to fully respect the IAEA recommendations concerning how to build a dose effect curve applicable to an acute exposure. This point must absolutely be considered for future intercomparisons as it has a very strong impact on the interpretability of the results and on the identification of improvement areas for a specific network. Additionally, evolution in the scoring criteria within a laboratory over the time elapsed between the calibration curve establishment and the present intercomparison may lead to additional uncertainty in the dose assessment. In fact, scorers are changing over the time so there is a need for periodical harmonization. Another important issue is how the participation in intercomparisons have modified the scoring criteria and so the dose-effect curve. This is particularly important if the dose-effect curves were produced prior to the harmonization work carried out within the RENEB network.

One more issue brought forward by this intercomparison was the lack of homogeneity in the calculation of the uncertainties associated to the doses reported by the participants. In fact, there are different generally accepted ways to estimate a dose and its associated uncertainties, as several calculation software programs are available (CABAS, Dose Estimate and Microsoft Excel-based spreadsheets). Considering that these tools do not implement the same methodologies to calculate uncertainties, it made it difficult to compare the raw reported values because they were not calculated in a homogenous manner. In the present study, the estimated doses and uncertainties initially sent by participants were calculated by each laboratory using their own methods. This led to a great heterogeneity in the reported values and in the reported curve coefficients, which further complicated their interpretation in the context of an intercomparison. For this reason, all dose estimates were re-calculated using the reported frequencies and their own dose-response curves using the method described by Merkle et al (1983) and mentioned in the IAEA manual (2011). For future ILCs, it seems essential to clearly define the methodology to be applied by the laboratories for the calculation of the dose and the associated uncertainties. To go further, the implementation of a single integrated and open-ended tool available to the participants seems to be relevant. This was the strategy adopted by the RENEB association, through the development of BiodoseTools, a software based on R with a Shiny interface (https://github.com/biodosetools-team/biodosetools).

## 5- CONCLUSION AND PERSPECTIVES

Standardization of chromosomal aberration scoring during the various European projects (MULTIBIODOSE and RENEB) has improved results of dose assessment in ILC exercises (Jaworska et al. 2015, Oestreicher et al. 2017). The present work demonstrates that harmonized and trained networks such as RENEB (in terms of chromosomal aberration scoring) obtain better results than a non-harmonized group. This is illustrated by the high level of satisfactory results obtained either in frequency or dose by L1-L20 when using classical intercomparison analysis tools, such as the Z-score and its associated decision thresholds. However, one cannot conclude that RENEB laboratories are fully harmonized, not only for those non-satisfactory results but also by the statistical tools used. These statistical tools appear to be limited and are not able to advantageously exploit the richness of results from large intercomparisons. At present, these tools do not allow a fine diagnosis of laboratory performance, neither do they serve as new avenues for improvement for the network of laboratories. For example, it would seem interesting to be able to easily discriminate results such as those obtained by the L5, L7 and L9, which intuitively do not seem equivalent, but are considered as such by looking at their Z- and U-scores. After this first stage of harmonization using these tools, the use of other approaches to test laboratory performance in future intercomparisons seems to be necessary. Solutions based on the bias-variance trade-off are currently being explored.

Another issue highlighted by this ILC is the question of the infinite diversity of dose-effect curves. It clearly appears that, despite the initial advantage of including the scoring specificities of each laboratory, the lack of recommendations and minimum criteria to evaluate the robustness of each laboratory's curve seems to be a negative point for the model of "a curve per lab". The construction of a robust curve is a long-term procedure, which should be part of a constant and dynamic evolution process in order to take into account the changes occurring over time in the laboratories, or the evolutions inherent to the process of harmonization of a network. In addition, the relevance of a dose-effect curve established 25 or 30 years ago by members who are no longer present in a given laboratory is questionable. By definition, the process of harmonization would generate a change in practices and may raise questions about the validity of a pre-existing dose-response curve. One of the main advantages of a large laboratory network

is its power in terms of data production. The present intercomparison generated the analysis of a striking 20,000 different metaphases per dose. When harmonization of practices is achieved, such a network could build an extremely robust dose-response curve in just 2 or 3 intercomparisons. This would also have the advantage of consolidating practices in terms of calculating coefficients and the associated uncertainties, making it a strategy that should be seriously considered in large.

Finally, and in the same spirit of unification, it seems relevant to develop tools that are better adapted to the collection and processing of results produced by the various participant laboratories. For the moment, this collection happens at a relatively small-scale (notably through the exchange of spreadsheet files). The coupling of tools such as BiodoseTools and web portals for collecting results seems to be particularly promising, both in terms of definition and application of the methodologies necessary for their processing (in particular, the calculation of uncertainties), but also in terms of the reliability associated with the traceability of results.

After a first harmonization phase lasting more than ten years (Kulka et al. 2017, Oestreicher et al. 2017, Gregoire et al. 2017), and even if there is still room for improvement, the level of harmonization reached by RENEB members definitely confirms the operational value of international networks of biological dosimetry laboratories, particularly in the case of large-scale radiological accidents.

ACKNOWLEDGEMENTS & DISCLAIMERS

One author's (WFB) efforts in this study was funded by AFRRI's intramural protocols RBB44313 and AFR-B4-4313. The author (WFB) wishes to thank Uma Subramanian and Dr David L. Bolduc for their contributions in this study. The opinions and assertions expressed herein are those of the author (WFB) and do not necessarily reflect the official policy or position of the Uniformed Services University of the Health Sciences or the United States Department of Defense. The research protocol was reviewed and approved by the USUHS IRB Committee in accordance with all Federal regulations governing the protection of humans in research.

DISCLOSURE STATEMENT

No conflict of interest was reported by the authors

BIBLIOGRAPHY


1. Ainsbury E, Livingston GK, Abbott MG, Moquet JE, Hone PA, Jenkins MS, Christensen DM, Lloyd DC and Rothkamm K. Interlaboratory variation in scoring dicentric chromosomes in a case of partial-body x-ray exposure: implications for biodosimetry networking and cytogenetic ''triage mode'' scoring. 2009. Radiat Res. 172:746–752.
2. Ainsbury EA and Lloyd DC. 2010. Dose estimation software for radiation biodosimetry. Health Phys. 98:290–295.
3. Ainsbury EA, Al-Hafidh J, Bajinskis A, Barnard S, Barquinero JF, Beinke C, de Gelder V, Gregoire E, Jaworska A, Lindholm C, Lloyd D, Moquet J, Nylund R, Oestreicher U, Roch-Lefévre S, Rothkamm K, Romm H, Scherthan H, Sommer S, Thierens H, Vandevoorde C, Vral A and Wojcik A. 2014 Feb. Inter- and intra-laboratory comparison of a multibiodosimetric approach to triage in a simulated, large scale radiation emergency. Int J Rad Biol. 90:193–202.
4. Bakkiam D, Bhavani M, Anantha Kumar AA, Sonwani S, Venkatachalam P, Sivasubramanian K and Venkatraman B. 2015. Dicentric assay: inter-laboratory comparison in Indian laboratories for routine and triage applications. Appl Radiat Isot. 99:77–85.
5. Bhavani M, Tamizh Selvan G, Kaur H, Adhikari JS, Vijayalakshmi J, Venkatachalam P, Chaudhury NK. 2014. Dicentric chromosome aberration analysis using giemsa and centromere specific fluorescence in-situ hybridization for biological dosimetry: An inter- and intra-laboratory comparison in Indian laboratories. Appl Radiat Isot. 92:85–90.
6. Beinke C, Oestreicher U, Riecke A, Kulka U, Meineke V, Romm H. Inter-laboratory comparison to validate the dicentric assay as a cytogenetic triage tool for medical management of radiation accidents. 2011. Radiat Meas. 46:929-935.
7. Beinke C, Barnard S, Boulay-Greene H, De Amicis A, De Sanctis S, Herodin F, Jones A, Kulka U, Lista F, Lloyd D, Martigne P, Moquet J, Oestreicher U, Romm H, Rothkamm K, Valente M, Meineke V, Braselmann H and Abend M. 2013. NATO dosimetry study Laboratory Intercomparison of the Dicentric Chromosome Analysis Assay. Radiat Res. 180:129–137.

8. Benjamini Y and Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat (Methodol). 57(1):289-300.
9. Berry AC. 1941. The accuracy of the Gaussian approximation to the sum of independent variates. Trans. Amer. Math. Soc. 49(1):122-136.
10. Christie DH, Chu MC and Carr Z. 2010. Global networking for biodosimetry laboratory capacity surge in radiation emergencies. Health Phys. 98:168–17.
11. Deperas J, Szluinska M, Deperas-Kaminska M, Edwards A, Lloyd D, Lindholm C, Romm H, Roy L, Moss R, Morand J, Wojcik A. 2007. CABAS: a freely available PC program for fitting calibration curves in chromosome aberration dosimetry. Radiat Protect Dosimetry. 124:115–123.
12. Di Giorgio M, Barquinero JF, Vallerga MB, Radl A, Taja MR, Seoane A, De Luca J, Stuck Oliveira M, Valdivia P, Garcıa Lima O, Lamadrid A, Gonzalez Mesa J, Romero Aguilera I, Mandina Cardoso T, Guerrero Carvajal YC, Arceo Maldonado C, Espinoza ME, Martınez-Lopez W, Mendez-Acuna L, Di Tomaso MV, Roy L, Lindholm C, Romm H, Guclu I and Lloyd D. 2011. Biological dosimetry intercomparison exercise: an evaluation of triage and routine mode results by robust methods. Radiat Res. 175:638–649.
13. Esseen CG. 1942. On the Liapunoff limit of error in the theory of probability. Arkiv for Matematik Astronomi och Fysik. A28 (9):1-19.
14. Garcia O, Di Giorgio M, Vallerga MB, Radl A, Taja MR, Seoane A, De Luca J, Stuck Oliveira M, Valdivia P, Lamadrid AI, Gonzalez JE, Romero I, Mandina T, Pantelias G, Terzoudi G, Guerrero-Carbajal C, Arceo Maldonado C, Espinoza M, Oliveros N, Martınez-Lopez W, Di Tomaso MV, Mendez-Acuna L, Puig R, Roy L and Barquinero JF. 2013. Interlaboratory comparison of dicentric chromosome assay using electronically transmitted images. Radiat Prot Dosimetry. 154(1):18–25.
15. Gregoire E, Kulka U, Barrios L, Ainsbury E, Bassinet C, Fattibene P, Oestreicher U, Pantelias, G, Terzoudi G, Trompier F, Voisin P, Vral A, Wojcik A and Roy L. 2017 Jan. The harmonization process to set up and maintain an operational biological dosimetry and physical retrospective dosimetry network: QA QM applied to the RENEB network. Int J Rad Biol. 93(1): 81-86.
16. International Atomic Energy Agency (IAEA). 2001. Manual 405 Vienna: IAEA
17. International Atomic Energy Agency (IAEA). 2011. Cytogenetic dosimetry: applications in preparedness for and response to radiation emergencies. Vienna: IAEA.

18. International Organization for Standardization (ISO) 13528: 2015. Statistical methods for use in proficiency testing by interlaboratory comparison.

19. International Organization for Standardization (ISO) 5725. 1998. Precision of test methods-determination of repeatability & reproducibility for a standard test method by interlaboratory tests..

20. International Organization for Standardization (ISO). 19238. 2014. Radiation protection-performance criteria for service laboratories performing biological dosimetry by cytogenetics. Geneva: ISO.

21. Jaworska A, Ainsbury EA, Fattibene P, Lindholm C, Oestreicher U, Rothkamm K, Romm H, Thierens H, Trompier F, Voisin P, Vral A, Woda C and Wojcik A. 2015 Apr. Operational guidance for radiation emergency response organizations in Europe for using biodosimetric tools developed in EU MULTIBIODOSE project. Radiat Protect Dosimetry. 164:1–5.

22. Kulka U, Ainsbury, EA, Atkinson M, Barquinero JF, Barrios L, Beinke C, Bognar G, Cucu A, Darroudi F, Fattibene P, Gil O, Gregoire E, Hadjidekova V, Haghdoost, Herranz R, Jaworska A, Lindholm C, M'kacher R, Möertl S, Montoro A, Moquet J, Moreno M, Ogbazghi A, Oestreicher U, Palitti F, Pantelias G, Popescu I, Prieto MJ, Romm H, Rothkamm K, Sabatier L, Sommer S, Terzoudi G, Testa A, Thierens H, Trompier F, Turai I, Vandesickel V, Vaz P, Voisin P, Vral A, Ugletveit F, Woda C and Wojcik A. Realising the European network of biodosimetry (RENEB). 2012. Radiat. Prot. Dosimetry. 151(4): 621-625.

23. Kulka U, Ainsbury, EA, Atkinson M, Barnard S, Smith R, Barquinero JF, Barrios L, Bassinet C, Beinke C, Cucu A, Darroudi F, Fattibene P, Bortolin F, Della Monaca S, Gil O, Gregoire E, Hadjidekova V, Haghdoost S, Hatzi V, Hempel W, Herranz R, Jaworska A, Lindholm C, Lumniczky K, M'kacher R, Möertl S, Montoro A, Moquet J, Moreno M, Noditi M, Ogbazghi A, Oestreicher U, Palitti F, Pantelias G, Popescu I, Prieto MJ, Roch-Lefévre S, Roessler U, Romm H, Rothkamm K, Sabatier L, Sebastia N, Sommer S, Terzoudi G, Testa A, Thierens H, Trompier F, Turai I, Vandevoorde C, Vaz P, Voisin P, Vral A, Ugletveit F, Wieser A, Woda C and Wojcik A. Realising the European network of biodosimetry: RENEB-status quo. 2015. Radiat. Prot. Dosimetry. 164:42–45.

24. Kulka U, Abend M, Ainsbury E, Badie C, Barquinero JF, Barrios L, Beinke C, Bortolin E, Cucu A, De Amicis A, Dominguez I, Fattibene P, Frovig AM, Gregoire E, Guogyte K, Hadjidekova V, Jaworska A, Kriehuber R, Lindholm C, Lloyd D, Lumniczky K, Lyng F, Meschini R, Mörtl S, Della

Monaca S, Monteiro Gil O, Montoro A, Moquet J, Moreno M, Oestreicher U, Palitti F, Pantelias G, Patrono C, Piqueret-Stephan L, Port M, Prieto MJ, Quintens R, Ricoul M, Romm H, Roy L, Sáfrány G, Sabatier L, Sebastià N, Sommer S, Terzoudi G, Testa A, Thierens H, Turai I, Trompier F, Valente M, Vaz P, Voisin P, Vral A, Woda C, Zafiropoulos D and Wojcik A. 2017. RENEB – Running the European Network of biological dosimetry and physical retrospective dosimetry. Int J Rad Biol. 93(1):2-14.

25. Liu JX, Pan Y, Ruan JL, Piao C, Su X. 2016. Intercomparison in cytogenetic dosimetry among 22 laboratories in China. Genome Integr. 7:6.
26. Lloyd DC, Edwards AA, Moquet JE, Guerrero-Carbajal YC. 2000. The role of cytogenetics in early triage of radiation casualties. Appl Radiat Isot. 52:1107-1112.
27. Merkle W. 1983. Statistical methods in regression and calibration analysis of chromosome aberration data. Radiat Environ Biophys. 21(3):217–233.
28. Miller SM, Ferrarotto CL, Vlahovich S, Wilkins RC, Boreham DR and Dolling J A. 2007. Canadian cytogenetic emergency network (CEN) for biological dosimetry following radiological/nuclear accidents. Int J Rad Biol. 83:471–477.
29. Oestreicher U, Samaga D, Ainsbury E, Antunes AC, Baeyens A, Barrios L, Beinke C, Beukes P, Blakely WF, Cucu A, De Amicis A, Depuydt J, De Sanctis S, Di Giorgio M, Dobos K, Dominguez I, Ngoc Duy P, Espinoza ME, Flegal FN, Figel M, Garcia O, Monteiro Gil O, Gregoire E, Guerrero-Carbajal C, Güçlü İ, Hadjidekova V, Hande P, Kulka U, Lemon J, Lindholm C, Lista F, Lumniczky K, Martinez-Lopez W, Maznyk N, Meschini R, M'kacher R, Montoro A, Moquet J, Moreno M, Noditi M, Pajic J, Radl A, Ricoul M, Romm H, Roy L, Sabatier L, Sebastià N, Slabbert J, Sommer S, Stuck Oliveira M, Subramanian U, Suto Y, Que T, Testa A, Terzoudi G, Vral A, Wilkins R, Yanti LY, Zafiropoulos D and Wojcik A. 2017. RENEB intercomparisons applying the conventional Dicentric Chromosome Assay (DCA). Int J Rad Biol. 93(1):20-29.
30. Pan Y, Ruan J, Gao G, Wu L, Piao C, and Liu J. 2019 Jan-Mar. Laboratory intercomparison of cytogenetic dosimetry among 38 laboratories in china. Dose-Response. 17(1):1-7.
31. Romm H, Wilkins RC, Coleman CN, Lillis-Hearne PK, Pellmar TC, Livingston GK, Awa AA, Jenkins MS, Yoshida MA, Oestreicher U and Prasanna PGS. 2011. Biological dosimetry by the triage dicentric chromosome assay: potential implications for treatment of acute radiation syndrome in radiological mass casualties. Radiat Res. 175:397–404.

32. Romm H, Ainsbury E, Bajinskis A, Barnard S, Barquinero JF, Beinke C, Puig-Casanovas R, Deperas-Kaminska M, Gregoire E, Kulka U, Oestreicher U, Lindholm C, Moquet J, Rothkamm K, Sommer S, Thierens H, Vral A, Vandersickel V, Wojcik A. 2014a. Web-based scoring of the dicentric assay, a collaborative biodosimetric scoring strategy for population triage in large scale radiation accidents. Radiat Environ Biophys. 53(2):241-254.

33. Romm H, Ainsbury E, Barnard S, Barrios L, Barquinero JF, Beinke C, Deperas M, Gregoire E, Koivistoinen A, Lindholm C, Moquet J, Oestreicher U, Puig R, Rothkamm K, Sommer S, Thierens H, Vandersickel V, Vral A, Wojcik A. 2014b June. Validation of semi-automatic scoring of dicentric chromosomes after simulation of three different irradiation scenarios, Health Phys. 106(6):764-71.

34. Romm H, Beinke C, Garcia O, Di Giorgio M, Gregoire E, Livingston G, Lloyd D, Martinez-Lopez W, Moquet JE, Sugarman SL, Wilkins RC and Ainsbury EA. 2016. A new cytogenetic biodosimetry image repository for the dicentric assay. Radiat Prot Dosimetry. 172(1-3): 192–200

35. Roy L, Buard V, Delbos M, Durand V, Paillole N, Gregoire E and Voisin P. 2004. International intercomparison for criticality dosimetry: the case of biological dosimetry. Radiat Prot Dosimetry. 110(1-4):471–476.

36. Shaffer JP. 1995. Multiple hypothesis testing. Annual Review of Psychology. 46:561-584.

37. Trompier F, Baumann M, Barrios L, Gregoire E, Abend M, Ainsbury E, Barnard S, Barquinero JF, Bautista JA, Brzozowska B, Perez-Calatayud J, De Angelis C, Domínguez I, Hadjidekova V, Kulka U, Mateos JC, Meschini R, Monteiro Gil O, Moquet J, Oestreicher U, Montoro Pastor A, Quintens R, Sebastià N, Sommer S, Stoyanov O, Thierens H, Terzoudi G, Villaescusa JI, Vral A, Wojcik A, Zafiropoulos D and Roy L. 2017 Jan. Investigation of the influence of calibration practices on cytogenetic laboratory performance for dose estimation. Int J Rad Biol. 93(1):118-126.

38. Vaurijoux A, Gruel G, Gregoire E, Roch-Lefevre S, Voisin Pa, Martin C, Voisin Ph, Roy L, Barquinero JF. 2015. Automatic dicentric scoring a real option to be used in biological dosimetry. Rad Emerg Med. 4:16-21.

39. Voisin P. Standards in biological dosimetry: A requirement to perform an appropriate dose assessment. 2015. Mutat Res. 793:115–122.

40. Wilkins R, Romm H, Kao TC, Awa AA, Yoshida MA, Livingston GK, Jenkins MS, Oestreicher U, Pellmar TC, and Prasanna PGS. 2008. Interlaboratory comparison of the dicentric chromosome assay for radiation biodosimetry in mass casualty events. Radiat Res. 169(5):551-560.

41. Wilkins RC, Romm H, Oestreicher U, Marro L, Yoshida M A, Suto Y, and Prasanna PGS. 2011 Sept. Biological dosimetry by the triage dicentric chromosome assay – further validation of international networking. Radiat Meas. 46(9): 923–928.

42. Wojcik A, Lloyd D, Romm H and Roy L, Biological dosimetry for triage of casualties in a large-scale radiological emergency: capacity of the EU member states. 2010. Radiat Prot Dosimetry. 138(4):397–401.

43. Yoshida MA, Hayata I, Tateno H, Tanaka K, Sonta S, Kodama S, Kodama Y and Sasaki MS. 2007. The chromosome network for biodosimetry in Japan. Radiat Meas. 42:1125–1127.

Figures and Tables Legends

**Table 1. Reported scoring values from each laboratory; dicentric (dic) and dicentric plus ring (dic+r).** In italics, reported values that did not include the cell distribution of aberrations. L1 to L20 are RENEB members and L21 to L38 are non-RENEB group. *L2 and L2b represent different kinds of staining (Giemsa (L2) or FISH coupled with pan-telomeric and pan-centromeric probes (L2b)). **L31 and L31b represent chromosomal aberration scoring (dicentrics (L31) and dicentrics plus rings (L31b)).

**Figure 1: Dicentric frequencies per metaphase for sample A (A), sample B (B) and sample C (C) from each of the participant laboratories.** Triangles represent the dicentric frequency per metaphase obtained by each laboratory for sample A (A), sample B (B) and sample C (C). The solid black line is the robust mean when laboratories reported the dicentric frequency per metaphase. Dashed lines mark the 95% confidence interval of the robust mean. Z-and U-scores were calculated with a Benjamini-Hochberg adjustment. Stars denote an unsatisfactory U-score and circles denote an unsatisfactory Z score. RENEB laboratories are within the grey rectangle.

**Table 2. Calibration curve coefficients of the participant laboratories. NA**: Not Available. These laboratories have no dose-effect curve. Some laboratories did not include standard deviations for the coefficients (± NA). ***L2 and L2b** represent different kinds of staining (Giemsa (L2) or FISH coupled with pan-telomeric and pan-centromeric probes (L2b)). ****L4/L5 and L4b/L5b** represent calibration curves based on different irradiation sources (gamma- (L4 and L5) and x-rays (L4b and L5b)). *****L31 and L31b** represent chromosomal aberration scoring (dicentrics (L31) and dicentrics plus rings (L31b)). L1 to L20 are RENEB members and L21 to L38 are non-RENEB group.

**Table 3: Source and dose rate used by the laboratories for their calibration curve.** * : L2 used Giemsa for dicentric chromosomes and centric rings staining, and L2b used TC-FISH for dicentric chromosomes and centric rings staining. ** : L31 scored only dicentric chromosomes to build its curve

and L31b scored dicentric chromosomes and centric rings to build its curve. L1 to L20 are RENEB members and L21 to L38 are non-RENEB group. **NP** : Not Provided

**Figure 2: Dose estimations sent by the participant laboratories for samples A, B, C.** Solid circles represent the dose estimation based on the total number of metaphases analyzed. Some laboratories sent two estimated doses for each sample, which are represented by open triangles. Error bars correspond to the reported 95% confidence interval. The horizontal line represents the delivered physical dose to the blood.

**Figure 3: Re-calculated doses by the IRSN based on Merckle's approach and using each laboratory's own curve coefficients for sample A (3A), sample B (3B) and sample C (3C).** Diamonds represent the average dose obtained and error bars correspond to the 95% confidence interval of the estimated dose. Values considered as unsatisfactory by the U-test are indicated with a star. Unsatisfactory results by the Z-test are indicated with a circle. RENEB laboratories are within the grey rectangle.

**Table 4: Laboratory ranking by Z-Score (A) and by U-Score (B). a** : CA Frequency : Frequency of chromosomal aberrations (dicentric chromosomes per cell)

**Table 5: Comparison of laboratory rankings between the Z-score obtained for dicentric frequency per metaphase and the Z-score obtained for assessed dose.** Rectangles show the laboratories whose rank changes the most between frequency and dose. L9 is highlighted by a solid line rectangle and L18 is highlighted by a dashed line rectangle. L36, L37 and L38 are not present in the dose column since they did not provide dose estimations. L4b, L5b and L31b are present only in the dose column because the dicentric frequencies are similar within the same laboratory (L4/L4b ; L5/L5b ; L31/L31b). **NA**: Not Available: The Z-score for L2b was not calculated because the staining technique (TC-FISH) was different from the rest (GIEMSA staining) and thus could not be compared using this test.

**Figure 4: Calibration curves of the Inter-Laboratory Comparison participants.** The horizontal line represents the frequency of 0.5 dicentric chromosomes or dicentrics + centric rings per metaphase

and the grey vertical lines indicate the mean estimated dose obtained with the two most distant curves. The dashed curve, indicated by an arrow, is that of the IAEA manual (IAEA 2011).

**Table 6: Comparison of robust values among laboratory categories.** x, s and CV correspond respectively to the calculated robust mean, robust standard deviation and robust coefficient of variation calculated.