Between-site reliability of startle prepulse inhibition across two early psychosis consortia

Kristin S. Cadenhead[a,b], Jean Addington[p], Tyrone D. Cannon[f], Barbara A. Cornblatt[h,i,j], Camilo de la Fuente-Sandoval[q], Dan H. Mathalon[c,d], Diana O. Perkins[k], Larry J. Seidman[l], Ming Tsuang[a], Elaine F. Walker[m], Scott W. Woods[g], Peter Bachman[e], Ayse Belger[k], Ricardo E. Carrión[h,i,j], Franc C.L. Donkers[k], Erica Duncan[o,n], Jason Johannesen[f], Pablo León-Ortiz[q], Gregory Light[a,b], Alejandra Mondragón[q], Margaret Niznikiewicz[l], Jason Nunag[a], Brian J. Roach[c,d], Rodolfo Solís-Vivanco[q] and The North American Prodromal Longitudinal Studies Consortium

Prepulse inhibition (PPI) and reactivity of the acoustic startle response are widely used biobehavioral markers in psychopathology research. Previous studies have demonstrated that PPI and startle reactivity exhibit substantial within-site stability; however, between-site stability has not been established. In two separate consortia investigating biomarkers of early psychosis, traveling participants studies were carried out as a part of quality assurance procedures to assess the fidelity of data across sites. In the North American Prodromal Longitudinal Studies (NAPLS) consortium, eight normal participants traveled to each of the eight NAPLS sites and were tested twice at each site on the startle PPI paradigm. In preparation for a binational study, 10 healthy participants were assessed twice in both San Diego and Mexico City. Intraclass correlations between and within sites were significant for PPI and startle response parameters, confirming the reliability of startle measures across sites in both consortia. There were between-site differences in startle magnitude in the NAPLS study that did not appear to be related to methods or equipment. In planning multisite studies, it is essential to institute quality assurance procedures early and establish between-site reliability to assure comparable data across sites.

NeuroReport 2013, **24**:626–630

Keywords: endophenotype, prepulse inhibition, reliability, startle

[a]Department of Psychiatry, University of California San Diego (UCSD), [b]Veteran's Affairs San Diego Healthcare System, La Jolla, [c]Department of Psychiatry, University of California San Francisco, [d]San Francisco VA Medical Center, San Francisco, [e]Department of Psychology, University of California Los Angeles (UCLA), Los Angeles, California, Departments of [f]Psychology, [g]Psychiatry, Yale University, New Haven, Connecticut, [h]Department of Psychiatry, The Zucker Hillside Hospital, New York, [i]Center for Psychiatric Neuroscience, The Feinstein Institute for Medical Research, Manhasset, [j]Department of Psychiatry and Molecular Medicine, Hofstra North Shore-LIJ School of Medicine, Hempstead, New York, [k]Department of Psychiatry, University of North Carolina (UNC), Chapel Hill, North Carolina, [l]Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School Harvard, Boston, Massachusetts, Departments of [m]Psychology, [n]Psychiatry, Emory University, [o]Atlanta Veteran's Affairs, Atlanta, Georgia, USA, [p]Department of Psychiatry, University of Calgary, Calgary, Alberta, Canada and [q]Department of Psychiatry, National Institute of Neurology and Neurosurgery (INNN), Mexico City, Mexico

Correspondence to Kristin S. Cadenhead, MD, Department of Psychiatry, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0810, USA
Tel: +619 543 6445; fax: +619 543 7315; e-mail: kcadenhead@ucsd.edu

Received 31 March 2013 **accepted** 20 May 2013

Introduction

Prepulse inhibition (PPI) and reactivity of the acoustic startle response are widely used translational biomarkers in psychopathological research. PPI is an index of sensorimotor gating and is used in animal and human studies to understand brain disorders such as schizophrenia and Tourette's disorder that are characterized by gating impairments in the neural substrates that underlie sensory information processing [1]. In the PPI paradigm, weak lead stimuli inhibit the startle response to intense, abrupt stimuli (acoustic, visual, tactile) [2]. PPI is typically reduced in individuals with schizophrenia [3], stable with repeated within-site testing [4–10], heritable [11,12], and associated with genes of relevance to psychosis [13,14], suggesting its utility as an endophenotype and as a vulnerability marker for psychosis risk [15].

An increasing emphasis in schizophrenia research has been in the area of early detection and intervention. The use of biobehavioral markers such as PPI in the study of the prodromal phase of psychosis provides a means of not only identifying individuals at greatest risk for psychosis but also understanding neurodevelopmental abnormalities early in the course of illness that can contribute to better informed treatment [16]. Although it is possible to use empirically derived criteria for a prodromal psychosis syndrome [17] to identify individuals at increased risk

0959-4965 © 2013 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/WNR.0b013e3283637845

of psychotic illness, the 2 year psychotic conversion rate is between 15–35% [18], making it difficult to recruit a sufficient number of participants at any one site. Therefore, multisite studies are essential to attain sufficient statistical power to investigate the prodromal phase of illness.

However, for biomarkers such as PPI to be useful in multisite studies that are needed to increase statistical power, facilitate the identification of disease risk, increase the odds of finding uncommon genetic variation, or identification of relevant subgroups, the measures need to be stable with repeated assessment and reliable across sites [19]. Because differences in testing conditions and procedures across sites can introduce uncontrolled variance in experimental measures, it is essential to understand potential site differences and control variation across sites as much as possible. Although multisite studies have investigated PPI [19], to our knowledge, there are no published reports of between-site reliability of startle measures using normal participants traveling between sites. This study investigated the within-site and between-site reliability of PPI and startle reactivity in two consortia designed to identify vulnerability markers in early psychosis: the North American Prodromal Longitudinal Studies (NAPLS) consortium and a University of California Institute for Mexico and the United States (UCMEXUS).

Materials and methods

Participants

Participants included (i) eight healthy individuals recruited from each of the eight NAPLS sites [Emory, Harvard, University of Calgary, University of California Los Angeles, University of California San Diego (UCSD), University of North Carolina, Yale, Zucker Hillside] (age 19–30 years, four men and four women) and (ii) 10 healthy individuals (age 28–38 years, four men and six women), recruited from the UCSD and the National Institute of Neurology and Neurosurgery (INNN) in Mexico City. All nine institutions received approval from their individual ethics committees for the study. Participants provided written informed consent after the procedures were fully explained. Participants were excluded if they had the following: any concomitant medical or neurological illness, current substance abuse or dependence (excluding nicotine), any Axis I disorders (per Structured Interview for DSM-IV), or positive family history of psychosis.

Acoustic startle paradigm

Equipment and procedures were identical at the eight NAPLS sites as well as between UCSD and INNN. Manuals with equipment setup, testing procedures, and instructions to participants were developed in English and Spanish (for INNN). A meeting was held in Boston October 2009 to train all NAPLS sites; UCSD staff visited the Mexico site in September 2009 to train INNN using the same procedures [20].

Participants were screened for hearing impairment (> 45 dB, 1000 Hz). Smokers were allowed to smoke up to 30 min before startle testing to avoid nicotine withdrawal or intoxication. A customized startle-stimulus generating system (Grace Design Model m902 Amplifier and Neurobehavioral Systems Presentation Software) developed by the UCSD site was used for all sites. The sound was calibrated at all sites using a Quest Technologies 210 Sound Level Meter (Oconomowoc, Wisconsin, USA) and a custom-made PPI calibration session to ensure 70 dB for background noise and 115 dB for extended length startle bursts at each of the sites. Neurophysiologic recordings at NAPLS sites were performed using identical Biosemi systems and recording software (Biosemi, Amsterdam, the Netherlands). For the UCMEXUS study, data were recorded using NeuroScan equipment and software (NuAmps Digital EEG Amplifier; NeuroScan Labs, Sterling, Virginia, USA). Electrodes (Ag/AgCl) were placed below and at the outer canthus of the right eye with resistances less than 10 kΩ [20]. Startle stimuli were presented binaurally through identical headphones (TDH-39P) at all sites. A 70 dB [A] broadband background noise was used with a pulse (115 dB [A], 40 ms noise burst) presented either alone or following [30, 60, or 120 ms interstimulus interval (ISI)] a prepulse (86 dB [A], 20 ms noise burst). The paradigm began with a 5-min acclimation period, then five pulse-alone stimuli followed by 30 trials consisting six trials each of the three prepulse conditions and 12 pulse-alone stimuli presented in a fixed, pseudorandom order. The paradigm ended with five more pulse-alone stimuli for a total of 40 trials. EMG activity for both consortia was analyzed at UCSD using Brain Vision Analyzer (Brain Vision LLC, Morrisville, North Carolina, USA) and high-pass filtered at 28 Hz at 12 dB/oct. Waveforms were smoothed using a 40 Hz 24 dB/oct low-pass filter. All trials were manually inspected for artifacts. Startle data were analyzed using wave-form averaging for each of the four different trial types within each block, after applying baseline correction and rectification of the data. The magnitude of the peak startle response (the highest point relative to baseline between 30 and 120 ms after onset of startle stimulus) was determined. All participants demonstrated a robust startle response to the first block of startle stimuli; however, participants who demonstrated a relative lack of startle stimulus elicited eye blink to the second block of startle stimuli in any test session were excluded per established methods [20]. The following startle measures were examined: (i) reactivity, or the mean magnitude of response to pulse-alone stimuli, and (ii) PPI, the percentage of change in startle magnitude to prepulse + pulse versus pulse-alone trials [(pulse–prepulse + pulse)/pulse) × 100]. The stability of the startle measures between and within sites was assessed using

intraclass correlations (ICC; random consistency model) and repeated measures analysis of variance (ANOVA) design. All participants were tested twice at each site and traveled to the other sites within 3 months for NAPLS and within 1 year (mean 5.5 months) for UCMEXUS. The order of testing was balanced across sites in both studies with a specified order for NAPLS participants starting with the home site and within the ten participants who were included in the UCMEXUS study (five at UCSD first, five at INNN first).

Results

As shown in Table 1, within-site ICCs of startle and PPI variables were significant across all reactivity and PPI conditions in both NAPLS and UCMEXUS (Table 1). Between-site analyses were similarly performed comparing time 1 to time 1 and time 2 to time 2 across sites. All but the 30-ms PPI ($P = 0.056$) condition for UCMEXUS were significant. Finally, within-site and between-site ICCs were calculated using both sessions at each of the sites and all were significant. In repeated measures ANOVA of PPI (NAPLS: 8 sites $\times$ 2 sessions $\times$ 3 ISIs; UCMEXUS: 2 sites $\times$ 2 sessions $\times$ 3 ISIs) there were no statistical site [NAPLS: $F(7,108) = 0.45$, NS; UCMEXUS: $F(1,9) = 0.51$, NS] or session [NAPLS: $F(1,108) = 0.72$, NS; UCMEXUS: $F(1,9) = 0.83$, NS] main or interaction effects supporting the within-site and between-site reliability. In contrast, a repeated measures ANOVA of startle reactivity (NAPLS: 8 sites $\times$ 2 times $\times$ 3 blocks; UCMEXUS: 2 sites $\times$ 2 times $\times$ 3 blocks) revealed a significant site effect [$F(7,28) = 2.46$, $P < 0.05$] for the NAPLS study due to one site having greater startle amplitude relative to the other sites (Fig. 1), but no session or interaction effects. When NAPLS site 4 was removed from the analysis, the significant site effect was no longer present [$F(6,30) = 1.97$, NS]. The site main effect for UCMEXUS [$F(1,9) = 0.48$, NS] was nonsignificant as were session and interaction effects.

Discussion

This is the first report of between-site reliability of PPI and startle reactivity measured with traveling participants. The present findings replicate previous studies that demonstrate within-site stability of startle measures in normal and schizophrenia spectrum participants [4–10] and extend these findings to demonstrate measurement comparability across laboratories in two separate multisite studies using two different types of equipment for neurophysiologic recording. It is likely that the standardization of equipment, protocols, training, analysis, and quality assurance procedures across sites contributed to the observed consistency of startle data.

Although startle reactivity was stable both within and between sites, significant site differences were observed in the NAPLS study driven by larger startle amplitude at one of the sites, prompting a review of equipment settings, stimulus calibration, ambient acoustic noise, electrical noise, placement of electrodes, participant instructions, and testing environment across sites. A decibel meter from UCSD was mailed to the site in question (site 4) to assure the loudness of the startle stimuli was accurate and consistent across sites. No methodological or equipment differences were identified. Individual participant data revealed that three participants had larger startle responses at site 4 (Fig. 2), accounting for the site differences. One participant with a large startle response was first exposed to the startle stimuli at site 4, perhaps accounting for the larger response. Since each participant began their travels at their home site, it is unlikely that order effects account for the observed differences. Thus, despite institution of careful quality assurance procedures, identical participants, methodology, and equipment, site differences still occur and need to be examined and controlled for in biomarker studies. Future analyses of NAPLS consortium data will continue to examine site differences in reactivity and site will be used as a between-participants factor.

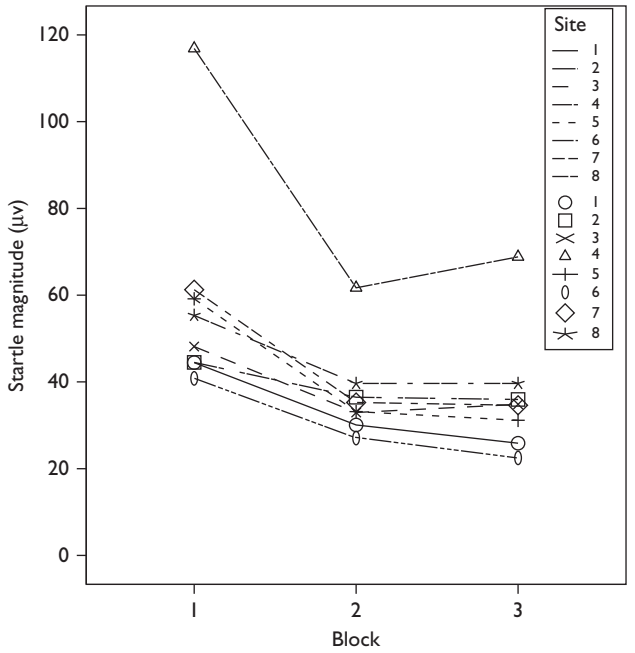
A limitation of the two studies is the relatively small sample size in each (NAPLS included eight participants tested at eight sites and UCMEXUS included 10 participants tested at two sites). The sample sizes, however, are consistent with the few traveling participants

Table 1 Intraclass correlations (random consistency model) of startle reactivity and prepulse inhibition within and between sites in the NAPLS and UCMEXUS studies

	NAPLS within-site ICC	NAPLS between-sites ICC	NAPLS within-site and between-site ICC	UCMEXUS within-site ICC	UCMEXUS between-site ICC	UCMEXUS within-site and between-site ICC
Startle reactivity						
Block 1	0.43***	0.46***	0.51***	0.85***	0.43*	0.60***
Block 2	0.80***	0.76***	0.79***	0.88***	0.52**	0.65***
%PPI						
30 ms PPI	0.67***	0.50***	0.57***	0.63***	0.31	0.38**
60 ms PPI	0.60***	0.57***	0.48***	0.67**	0.80***	0.52***
120 ms PPI	0.68***	0.73***	0.78***	0.59***	0.80***	0.69***

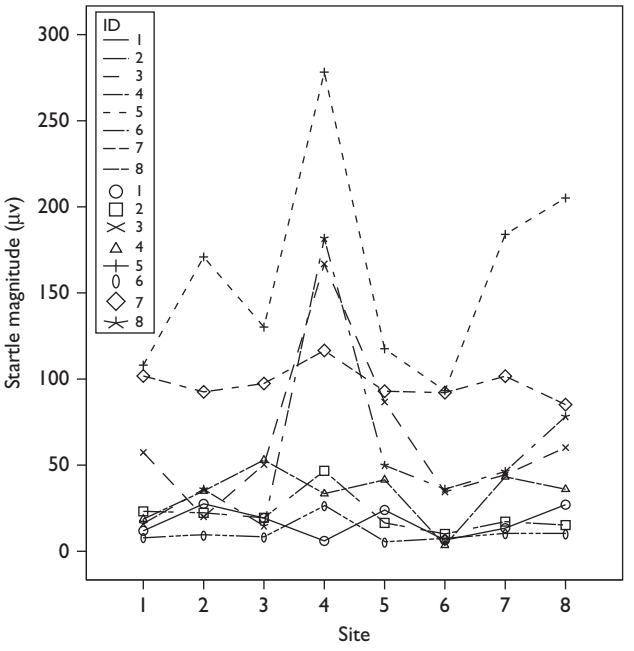
ICC, intraclass correlations; NAPLS, North American Prodromal Longitudinal Studies; PPI, prepulse inhibition; UCMEXUS, University of California Institute for Mexico and the United States.
*$P<0.05$.
**$P<0.01$.
***$P<0.001$.

Fig. 1



Site differences in startle reactivity were evident in the North American Prodromal Longitudinal Studies consortium. Data represents estimated marginal means that collapsed two test sessions at each site.

Fig. 2



Individual traveling participants data from the North American Prodromal Longitudinal Studies consortium demonstrates that three participants accounted for the observed site differences. Data represents estimated marginal means that collapse both test sessions and three blocks of startle magnitude in response to pulse-alone stimuli.

studies performed to establish reliability of neuroimaging measures across sites [21,22]. Although future between-site biomarker reliability studies should ideally use more participants, sending multiple participants to different cities and countries for multiple testing sessions obviously presents financial and logistical challenges.

Conclusion

In planning multisite biomarker studies, it is essential to institute standardized quality assurance procedures before data collection of targeted research samples. The use of identical equipment, training, and similar testing environments appears to be useful to minimize sources of cross-site variance in electrophysiological studies. The observed reliability of startle measures across laboratories provides support for the utility of these measures as biomarkers and endophenotypes in large multisite studies. Investigation and statistical control of potential site differences is essential in any multisite biomarker study.

Acknowledgements

The authors thank Kathleen Shafer, Daniel Roman, Clara Robles, Sylvana Stefano, Rafael Favila, Aaron Peterson, Sona Sandhu, Anna Evans, Kathleen Monforton, Brenton Roman, Joshua Kenney for their technical assistance with the project.

This research was supported by a collaborative grant from UCMEXUS–CONACYT (K.S.C. and C.d.l.F.-S., co-PIs), National Institutes of Health (NIH)-supported grants U01 MH081944 to K.S.C., U01 MH082022 to S.W.W., U01MH081984 to J.A., U01 MH081902 to T.D.C., UO1 MH081857 to B.A.C., U01MH082004 to D.O.P., U01 MH081928 to L.J.S., and U01MH081988 to E.F.W.

Conflicts of interest

Barbara A. Cornblatt has been a consultant for Hoffman La Roche and received royalties for the CPT-IP. Jean Addington has been a consultant for Hoffman La Roche. Diana O. Perkins is on the Advisory Board for Sunovion DSMB, Genentech CNS, Genentech Mosaic Registry and a Consultant for Telesage. Scott W. Woods has been a Consultant for Merck. Camilo de la Fuente-Sandoval has served as consultant and/or speaker for IMS Health, Carnot Laboratories, Eli Lilly, and Janssen. For the remaining authors there are no conflicts of interest.

References

1 Braff DL. Prepulse inhibition of the startle reflex: a window on the brain in schizophrenia. *Curr Top Behav Neurosci* 2010; **4**:349–371.
2 Graham FK. Presidential address, 1974. The more or less startling effects of weak prestimulation. *Psychophysiology* 1975; **12**:238–248.
3 Braff DL, Grillon C, Geyer MA. Gating and habituation of the startle reflex in schizophrenic patients. *Arch Gen Psychiatry* 1992; **49**:206–215.
4 Cadenhead KS, Carasso B, Swerdlow NR, Geyer MA, Braff DL. Prepulse inhibition and habituation of the startle response are stable neurobiological measures in a normal male population. *Biol Psychiatry* 1999; **45**:360–364.

5 Abel K, Waikar M, Pedro B, Hemsley D, Geyer M. Repeated testing of prepulse inhibition and habituation of the startle reflex: a study in healthy human controls. *J Psychopharmacol* 1998; **12**:330–337.

6 Schwarzkopf SB, McCoy L, Smith DA, Boutros NN. Test–retest reliability of prepulse inhibition of the acoustic startle response. *Biol Psychiatry* 1993; **34**:896–900.

7 Ludewig K, Geyer MA, Etzensberger M, Vollenweider FX. Stability of the acoustic startle reflex, prepulse inhibition, and habituation in schizophrenia. *Schizophr Res* 2002; **55**:129–137.

8 Quednow BB, Kuhn KU, Beckmann K, Westheide J, Maier W, Wagner M. Attenuation of the prepulse inhibition of the acoustic startle response within and between sessions. *Biol Psychol* 2006; **71**:256–263.

9 Flaten MA. Test–retest reliability of the somatosensory blink reflex and its inhibition. *Int J Psychophysiol* 2002; **45**:261–265.

10 Light GA, Swerdlow NR, Rissling AJ, Radant A, Sugar CA, Sprock J, *et al*. Characterization of neurophysiologic and neurocognitive biomarkers for use in genomic and clinical outcome studies of schizophrenia. *PLoS One* 2012; **7**:e39434.

11 Greenwood TA, Braff DL, Light GA, Cadenhead KS, Calkins ME, Dobie DJ, *et al*. Initial heritability analyses of endophenotypic measures for schizophrenia: the consortium on the genetics of schizophrenia. *Arch Gen Psychiatry* 2007; **64**:1242–1250.

12 Hasenkamp W, Epstein MP, Green A, Wilcox L, Boshoven W, Lewison B, Duncan E. Heritability of acoustic startle magnitude, prepulse inhibition, and startle latency in schizophrenia and control families. *Psychiatry Res* 2010; **178**:236–243.

13 Greenwood TA, Lazzeroni LC, Murray SS, Cadenhead KS, Calkins ME, Dobie DJ, *et al*. Analysis of 94 Candidate Genes and 12 Endophenotypes for Schizophrenia From the Consortium on the Genetics of Schizophrenia. *Am J Psychiatry* 2011; **168**:930–946.

14 Greenwood TA, Light GA, Swerdlow NR, Radant AD, Braff DL. Association analysis of 94 candidate genes and schizophrenia-related endophenotypes. *PLoS One* 2012; **7**:e29630.

15 Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* 2003; **160**:636–645.

16 Cadenhead KS. Vulnerability markers in the schizophrenia spectrum: implications for phenomenology, genetics, and the identification of the schizophrenia prodrome. *Psychiatr Clin North Am* 2002; **25**:837–853.

17 Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Cannon T, Ventura J, *et al*. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. *Schizophr Bull* 2003; **29**: 703–715.

18 Cannon TD, Cadenhead K, Cornblatt B, Woods SW, Addington J, Walker E, *et al*. Prediction of psychosis in youth at high clinical risk: a multisite longitudinal study in North America. *Arch Gen Psychiatry* 2008; **65**:28–37.

19 Swerdlow NR, Sprock J, Light GA, Cadenhead K, Calkins ME, Dobie DJ, *et al*. Multi-site studies of acoustic startle and prepulse inhibition in humans: initial experience and methodological considerations based on studies by the Consortium on the Genetics of Schizophrenia. *Schizophr Res* 2007; **92**:237–251.

20 Cadenhead KS. Startle reactivity and prepulse inhibition in prodromal and early psychosis: effects of age, antipsychotics, tobacco and cannabis in a vulnerable population. *Psychiatry Res* 2011; **188**:208–216.

21 Friedman L, Stern H, Brown GG, Mathalon DH, Turner J, Glover GH, *et al*. Test–retest and between-site reliability in a multicenter fMRI study. *Hum Brain Mapp* 2008; **29**:958–972.

22 Brown GG, Mathalon DH, Stern H, Ford J, Mueller B, Greve DN, *et al*. Multisite reliability of cognitive BOLD data. *Neuroimage* 2011; **54**: 2163–2175.