Published in final edited form as:

Schizophr Res. 2007 May ; 92(1-3): 237–251.

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

Neal R. Swerdlow¹, Joyce Sprock¹, Gregory A. Light¹, Kristin Cadenhead¹, Monica E. Calkins², Dorcas J. Dobie³, Robert Freedman⁴, Michael F. Green^{5,6}, Tiffany A. Greenwood¹, Raquel E. Gur², Jim Mintz⁵, Ann Olincy⁴, Keith H. Nuechterlein⁵, Allen D. Radant³, Nicholas J. Schork¹, Larry J. Seidman⁷, Larry J. Siever⁸, Jeremy M. Silverman⁸, William S. Stone⁷, Debbie W. Tsuang³, Ming T. Tsuang^{1,7}, Bruce I. Turetsky², and David L. Braff¹

1 Department of Psychiatry, University of California San Diego, San Diego, California, USA.

2 Department of Psychiatry, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

3 Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, and VA Puget Sound Health Care System, Seattle, Washington, USA, and VISN20 MIRECC.

4 Department of Psychiatry, University of Colorado Health Sciences Center, Denver, Colorado, USA.

5 Department of Psychiatry and Biobehavioral Sciences, Geffen School of Medicine at University of California Los Angeles, Los Angeles, California, USA.

Corresponding Author: Neal R. Swerdlow, M.D., Ph.D. Department of Psychiatry UCSD School of Medicine 9500 Gilman Dr. La Jolla, CA 92093-0804 nswerdlow@ucsd.edu

General inquiries related to COGS: David L. Braff, M.D. dbraff@ucsd.edu

Contributors: Neal R. Swerdlow contributed to study design, data analysis, quality assurance and interpretation, and wrote the first version of this manuscript. Joyce Sprock contributed to study design, data analysis, quality assurance and interpretation, and provided editorial assistance with this manuscript. The following authors contributed to study design and data interpretation, and provided editorial assistance with this manuscript: David L. Braff, Kristin Cadenhead, Monica E. Calkins, Dorcas J. Dobie, Michael F. Green, Raquel E. Gur, Jim Mintz, Keith H. Nuechterlein, Ann Olincy, Gregory A. Light, Allen D. Radant, Larry J. Seidman, Jeremy M. Silverman, William S. Stone, Debby W. Tsuang and Bruce I. Turetsky. Larry J. Siever contributed to study design and data interpretation. Robert Freedman, Nik Schork and Ming Tsuang contributed to study design. Tiffany A. Greenwood contributed to data interpretation.

Conflict of Interest: No authors had a conflict of interest related to the contents of this manuscript.

Test equipment for this multi-site study included:

1. System for stimulus programming and delivery, and response acquisition:

San Diego Instruments EMG-SR Startle System SR-Lab (San Diego, CA); includes EMG-SR software, PC interface board, Stimulus Module, all cables and connectors;

Headphones for startle stimulus delivery (Maico Model TDH-39-P; Minneapolis, MN);

Stimulus calibration: Quest Model 2700-10 dB Meter (Quest Electronics, Oconomowoc, WI), QC-10 sound calibrator and calibrator adapter ring (#056-990), adapter ring (#58-928), metal EC9A 6 cc earphone coupler and weight, W-440.

2. Equipment and supplies for electrode preparation and application:

Electrodes - In Vivo Metrics (Healdsburg, CA) model #E220X-LS 4 mm, 40 inch lead, gray;

Electrode washers - In Vivo Metrics, model #E401;

Gel - Electro-Cap International, Inc electrode gel, model E9 (Eaton, OH);

Impedance meter: UFI (Morro Bay, CA) model 1089 MKIII CHECKTRODE;

Skin preparation: Kendall Curity Gauze Sponges, 12 ply- 4 in × 3 in, USP Type VII Gauze and rubbing alcohol or PDI Electrode Prep Pads (with pumice & rubbing alcohol).

3. Other:

Electrical noise measurements - Sypris Triaxial ELF Magnetic Field Meter (Model 4080; Orlando, FL).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

6 VA Greater Los Angeles Healthcare System

7 Massachusetts Mental Health Center Public Psychiatry Division of the Beth Israel Deaconess Medical Center, Harvard Medical School Department of Psychiatry, Boston, Massachusetts, and Harvard Institute of Psychiatric Epidemiology and Genetics, Boston, Massachusetts, USA.

8 Department of Psychiatry, The Mount Sinai School of Medicine, New York, New York, USA, and James J. Peters VA Medical Center and VISN3 MIRECC.

Abstract

Background: Startle and its inhibition by weak lead stimuli (“prepulse inhibition”: PPI) are studied to understand the neurobiology of information processing in patients and community comparison subjects (CCS). PPI has a strong genetic basis in infrahumans, and there is evidence for its heritability, stability and reliability in humans. PPI has gained increasing use as an endophenotype to identify vulnerability genes for brain disorders, including schizophrenia. Genetic studies now often employ multiple, geographically dispersed test sites to accommodate the need for large and complex study samples. Here, we assessed the feasibility of using PPI in multi-site studies.

Methods: Within a 7-site investigation with multiple measures, the Consortium on the Genetics of Schizophrenia conducted a methodological study of acoustic startle and PPI in CCS. Methods were manualized, videotaped and standardized across sites with intensive in-person training sessions. Equipment was acquired and programmed at the “PPI site” (UCSD), and stringent quality assurance (QA) procedures were used. Testing was completed on 196 CCS over 2.5 years, with 5 primary startle dependent measures: eyeblink startle magnitude, habituation, peak latency, latency facilitation and PPI.

Results: Analyses identified significant variability across sites in some but not all primary measures, and determined factors both within the testing process and subject characteristics that influenced a number of test measures. QA procedures also identified non-standardized practices with respect to testing methods and procedural “drift”, which may be particularly relevant to multi-site studies using these measures.

Conclusion: With thorough oversight and QA procedures, measures of acoustic startle PPI can be acquired reliably across multiple testing sites. Nonetheless, even among sites with substantial expertise in utilizing psychophysiological measures, multi-site studies using startle and PPI as dependent measures require careful attention to methodological procedures.

Keywords

endophenotype; prepulse inhibition; schizophrenia; sex differences; startle

1. Introduction

Acoustic startle and its inhibition by weak lead stimuli (“prepulse inhibition”: PPI) are quantified easily via electromyographic (EMG) measures of the blink reflex in normal and disordered human populations (Graham 1975; cf. Braff et al. 2001b). Neurobiological substrates of automatic, uninstructed PPI have been elucidated in humans and non-human mammalian species (Kumari et al. 2003; Postma et al. 2006; cf. Swerdlow et al. 2000, 2001a), and quantitative trait loci associated with PPI have been identified in rodents (Jooper et al. 2002). Preclinical studies have demonstrated that strain differences in PPI reflect genetic rather than epigenetic influences (Francis et al. 2003; Swerdlow et al. 2004). When measured over intervals of several weeks, PPI is stable (Cadenhead et al. 1999), and some evidence suggests that it is a heritable phenotype in humans (Anokhin et al. 2003). These features underscore the importance of PPI deficits in certain complex neuropsychiatric disorders with suspected or

identified genetic origins, including schizophrenia (cf. Braff et al. 2001b; Braff et al. 1978, 1992, 1999, 2001a, 2005; Leumann et al. 2002; Swerdlow et al. 2006; Weike et al. 2000) and Tourette Syndrome (Castellanos et al. 1996; Swerdlow et al. 2001b). For these reasons, there has been increasing interest in the use of PPI as a quantitative physiological phenotype, or endophenotype, to identify vulnerability genes for these disorders (cf. Braff and Freedman 2002; Anokhin et al. 2003; Cadenhead et al. 2000; Gottesman and Gould 2003; Kumari et al. 2005; Braff and Light 2005; Braff et al. 2006; Turetsky et al. 2006).

Genetic studies of complex quantitative behavioral disorders often require multiple data collection sites to ascertain large samples or family cohorts. Such geographically diverse studies present challenges for the use of endophenotypes, because differences in methodologies or test conditions across sites can introduce uncontrolled variance into the experimental measures. This variance could potentially obscure detection of physiological signals used to identify vulnerability genes. Thus, before PPI can be used with full confidence as an endophenotype in studies of complex behavioral disorders, it will be important to determine whether PPI can be reliably acquired at multiple test sites, using uniform techniques. This issue is most easily examined in non-patient populations, since in this case, site differences in patient characteristics, symptoms, medications and other clinical variables do not impact the findings. The present study reports the initial PPI data gathered from non-patient community comparison subjects (CCS) in a multisite collaborative study, the Consortium on the Genetics of Schizophrenia (COGS) (cf. Calkins et al. 2006). Subjects completed standardized neurocognitive and neurophysiological tests, including tests of the acoustic startle response and PPI at seven sites, all of which used identical equipment, techniques, training and subject selection criteria. All data were analyzed at a central PPI Quality Assurance site (UCSD: see Figure 1), with substantial experience in startle waveform analysis. We hypothesized that we would observe comparable levels of acoustic startle and PPI across sites, thereby confirming the suitability of these measures for multisite genetic studies. Differences in subject characteristics across sites were also assessed, as such differences might be expected in multisite studies, and might contribute to variance that impacts the reliability of startle measures and PPI.

2. Methods

2.1 Participants

Participants were recruited as part of the COGS project and tested at seven geographically dispersed, University affiliated sites (in alphabetical order): Harvard University, Mount Sinai School of Medicine, University of California at Los Angeles, University of California at San Diego, University of Colorado, University of Pennsylvania and University of Washington. Participants were 18-65 years old and fluent in English. Only data from CCS are described in this report. At each site, medically healthy CCS were recruited directly through flyers, print, and electronic media. CCS were excluded from the study for: electroconvulsive therapy in the past 6 months, positive illicit drug or alcohol screen, diagnosis of substance abuse disorder in past 30 days or substance dependence in past 6 months, estimated IQ lower than 70, history of significant head injury (including any of: loss of consciousness > 1 min, post-concussive syndrome > 1 week, or abnormal brain imaging or electroencephalography after event), seizure disorder, and other ocular, auditory, neurological or major systemic medical problems, a personal history of cluster A personality disorder or psychosis, or a family history of psychosis in a first- or second-degree relative. Local IRB boards of each testing site approved the study and all participants provided signed informed consent before participation in the study procedures.

All participants underwent standardized diagnostic and clinical assessments by diagnosticians trained according to a standardized procedure, including a modified version of the Diagnostic

Interview for Genetic Studies (DIGS; (Nurnberger et al., 1994), Scale for the Assessment of Negative Symptoms (SANS; Andreason 1984a) and Positive Symptoms (SAPS; Andreason 1984b), and a review of relevant medical records. Each subject was assigned DSM-IV best estimate final diagnoses through a consensus process that included review by at least two faculty level clinicians.

Sample characteristics are seen in Table 1. Of the 196 CCS who underwent valid testing during 30 months of data collection, 23 (11.7 %) were excluded from analysis based on established criteria for unacceptably low startle magnitude (see Figure 2). One participant was excluded for a positive saliva alcohol measurement, one for a current diagnosis of alcohol dependence, and one for a current diagnosis of alcohol abuse. The final study sample included 170 participants. A complete description of clinical and neurocognitive instruments used to characterize participants is found in a published report (Calkins et al. 2006).

Study design—Startle testing was administered as part of the COGS research protocol consisting of 4 hours of clinical assessment and 6 hours of neurophysiological and neuropsychological testing. The COGS neurophysiological and neuropsychological tasks were presented in one of two standardized orders, and a brief rest period was allowed between tasks (Table 2). Dependent measure scoring was conducted blind to diagnostic group membership.

2.2 Test Procedures

Startle testing was initiated after completion of a specific set of diagnostic or experimental measures (Table 2). In 46 subjects, testing was divided over two days, but the test sequence was maintained.

A list of test equipment is provided as a footnote to this report. Audiometric screening excluded hearing impairment (exclusion for threshold >40 dB at 1000 Hz). For startle testing, subjects sat in a recliner chair in a sound-attenuated room. Methods followed previous reports (e.g. Braff et al. 2001a,2005). Two 4 mm Ag/AgCl electrodes were positioned below and lateral to each eye over orbicularis oculi, with a ground electrode behind the left ear. Electrode resistances were <10 kOhm. The eye-blink component of the acoustic startle response was measured using an EMG startle system that recorded 250 1-ms epochs, starting with startle stimulus onset. Recorded EMG activity was band-pass filtered (100-1000 Hz). Amplification was set at 10,000 (a 0.25 mV signal triggered a 2.5V amplifier output). Blink scoring parameters are based on criteria established by Graham (1975), described previously (Braff et al. 1992). A square wave calibrator established sensitivity (1.31 μ V/digital unit). A 60-Hz notch filter was used to reduce interference. Stimuli were presented binaurally through headphones. Sound levels were calibrated monthly. Ambient room noise was measured using an artificial ear coupler on the headphones. Electrical noise in the test room was measured using magnetic field meter.

Cross-site Training and Quality Assurance—Laboratory staff and key faculty members from each site participated in a two-day in-person training session at UCSD prior to the study's initiation. Staff also received annual, in-person, refresher training on the startle testing procedures. A comprehensive manual describing the equipment, software, subject placement, scripts for interacting with participants and task administration was written and distributed to all sites (available on request), as were videotapes of all testing procedures. Ongoing quality assurance consisted of data quality review, biweekly conference calls and consultation with the central site. In addition, the project administrator visited each site annually to review procedures, inspect equipment, and undergo startle testing (Calkins et al. 2006).

2.3 Data analysis (Figure 1)

Data were scored using San Diego's Instrument's SRREDNEW program, which generates baseline EMG levels during the 20 ms epoch after startle stimulus delivery, onset and peak response latency (ms) and response amplitude (digital units). Software parameters by which voluntary and spontaneous blinks were recognized and excluded were derived from published criteria (Braff et al 1978; Geyer & Braff 1982; Graham 1975). Onset latency was defined by an algorithm based on the initial deviation from baseline EMG levels, and peak latency was defined as the point of maximal amplitude within 100 ms from startle stimulus onset. To exclude non-startle-induced blinks, rectified waveforms were automatically flagged for visual inspection when the onset and peak latencies differed by more than 100 ms. Responses were also assigned an error code when baseline values shifted by more than 10 units. Exclusion of all but stimulus-elicited eye blinks precludes contamination of the measures from group-related differences and rates of spontaneous blinks, which may be abnormal in patient populations.

Each trial was visually inspected for spontaneous and voluntary blinks, noise created by movement artifact and on-going EMG activity occurring at the onset of the stimulus. Trials with artifact were excluded from analysis on a trial-by-trial basis (<4% total trial exclusion). The SRREDNEW program automatically selected the peak of the startle response waveform within the 250 ms post-stimulus window for analysis, and this selection was monitored, blind to subject identity. Mean startle latencies and amplitudes were generated for each trial type within each block for the right and left eye for each participant. The QA site then uploaded these averages to the central data website, along with comments on the validity and quality of the data.

2.4 Statistics

Demographic differences across test sites were analyzed by 1-way ANOVAs for continuous variables and chi-square test for categorical variables. For experimental measures (startle magnitude, habituation, latency and PPI), repeated measures ANOVAs with Greenhouse-Geisser corrections and Fisher's PLSD post-hoc comparisons were performed with test site and sex as between-subject factors. The ratio of the mean magnitude of startle with a prepulse to the amplitude of startle magnitude without a prepulse was determined. %PPI is equal to 1 minus this ratio, expressed as a percentage, i.e. $100 \times (1 - (\text{magnitude of startle to pulse preceded by prepulse}) / \text{magnitude of startle to pulse without a preceding prepulse})$. For %PPI, within-subject factors were block, prepulse interval and eye side. Analyses also compared measures of onset and peak reflex latency, latency facilitation (latency reduction on trials with a prepulse followed by pulse vs. pulse alone trials) and reflex habituation (startle magnitude reduction in trial block 4 vs. 1 calculated as both difference scores and % reduction) across sites. Additional "secondary" analyses examined the impact of specific testing features (e.g. test order, time of day, one vs. two test days) and participant characteristics (e.g. age, smoking behavior) on the key dependent measures, particularly where there were known inter-site sample (e.g. different sex distributions) or methodological differences (e.g. differences in test order). Because patterns of onset and peak latency were comparable for most analyses, only peak latency values are reported for secondary analyses. Correlations were assessed via simple regression, or when using a single value per site (e.g. ambient noise level), via Spearman Rank analyses. Alpha for all comparisons was 0.05. Effect sizes (Cohen's d (Cohen 1988)) are reported where appropriate.

3. Results

3.1 Sample characteristics across sites

Recruitment and ascertainment practices across sites yielded samples comparable in some but not all characteristics (Table 1). ANOVAs revealed significant differences in the Wide Range

Achievement Test, Third Edition (WRAT3) reading subscale scores across sites ($p=0.03$); Chi-square analyses revealed significant differences in CCS sex distributions across sites (% female, maximum vs. minimum = 75% vs. 29%; $p<0.01$), current Axis I non-schizophrenia diagnoses (% with current diagnoses, maximum vs. minimum = 22% vs. 0%, $p<0.05$) and ethnic representation (% non-Caucasian, maximum vs. minimum = 50% vs. 4%; $p<0.05$), and trends for differences in the proportion of CCS with lifetime Axis I diagnoses (% ever diagnosis, maximum vs. minimum = 50% vs. 12.5%, $p<0.1$), and the proportion of CCS that are currently using psychotropic medications (% using psychotropic medication, maximum vs. minimum = 16.7% vs. 0%; $p<0.2$).

Axis I diagnoses were identified in 45 (26.5%) of the CCS. Of these, 29 (17.1%) were diagnosed with a non-psychotic affective disorder (Major Depressive Disorder or Dysthymia), and 23 (13.5%) were diagnosed with a substance-related disorder (Dependence or Abuse). No CCS carried diagnoses known to be associated with reduced levels of PPI. Medications were used by 22 (12.9%) of the final sample; of these, most (13; 7.6%) were taking herbal supplements or non-prescription medications, and 8 (4.7%) were taking antidepressants.

3.2 Methodological differences across sites

Despite significant efforts towards cross-site standardization of methods (see above and Calkins et al. (2006)), QA efforts detected a number of differences across sites in the experimental conditions and methodologies related to startle testing (Table 3). These differences related to the placement of EMG electrodes, EMG channel configurations, subject instructions, levels of ambient acoustic noise (range: 34-45 dB(A)) and electrical noise (range = 0.1-1.2 gauss) in the test room, test order ($p<0.001$; % sample tested in order “A”, maximum vs. minimum = 94% vs. 35%) and use of “split test” designs (testing over more than one day) ($p<0.001$; % sample tested over > 1 day, maximum vs. minimum = 90% vs. 0%). Differences across sites were also detected in EMG electrode impedance ($p<0.0001$; range (mean/site) = 5.16-9.10 k Ω) and hearing threshold ($p<0.0001$; range (mean/site) = 7.00-20.31 dB(A)). Despite these differences, all mean site values for electrode impedance and hearing threshold fell within designated acceptable ranges.

3.3 Startle measures across sites

Primary analyses compared four major startle characteristics across sites. Secondary analyses were then used to identify specific predicted patterns of startle characteristics, and to examine any impact of inter-test site differences on these measures. Of the sample characteristics and methodological variables found to differ across sites (above), WRAT3 Reading scores, hearing threshold, electrode impedance, participant sex and ethnicity, test order and use of split testing could be assessed as potential contributors to variability in startle measures.

3.3.1. Startle magnitude (Figure 3)—ANOVA of startle magnitude on pulse alone (PA) trials during blocks 2 and 3, when PPI was assessed, revealed no significant main effect of test site ($F=1.18$, df 6, 163, ns), a significant effect of trial block (2 vs. 3: $F=168.03$, df 1,163, $p<0.0001$), and no significant interaction of site $\times$ block ($F<1$). There were no significant effects of eye side ($F<1$), or other 2- or 3-way interactions. Across the 7 test sites, mean startle magnitude (SD) on PA trials during blocks 2 and 3 ranged from 43.0 (28.0) units to 78.7 (37.3) units (56.3 (36.7) - 103.1 (48.9) μV). It is possible that this range of startle magnitude values across sites reflected the impact of any one of several identified cross-site differences in participant characteristics or methodologies. However, startle magnitude did not correlate significantly with either hearing threshold ($r=0.04$) or electrode impedance ($r=0.04$), suggesting that site differences in these measures did not contribute substantially to the range of startle values. Startle magnitude differed based on ethnicity, with Caucasians showing significantly greater magnitude than non-Caucasian participants ($F=9.73$, df 1,133, $p<0.003$),

consistent with previous reports (Swerdlow et al. 2005). Because site C used predominantly one test order, and split-testing was never done at site A but always done at site D, ANOVA was performed without a “site” factor, and detected no significant effects of sex, test order or split testing on startle magnitude, and no significant interactions, suggesting that these factors did not contribute significantly to the overall range of startle values. There was no significant correlation between mean startle magnitude and ambient noise levels at each site ($r_s=0.40$, $p=0.33$). Thus, none of the methodological variables that differed across sites independently accounted for the observed site variability in startle magnitude.

Analysis of startle magnitude on no-stim (blank) trials revealed no significant effects of test site, trial block, eye side or sex (all F 's<1), or any significant 2, 3 or 4-way interactions.

3.3.2. Startle habituation (Figure 4)—Startle magnitude normally declines with repeated stimulus presentation, and the amount of reflex habituation was assessed by comparing reflex magnitudes in trial blocks 1 vs. 4. ANOVA of startle magnitude during blocks 1 and 4 revealed a trend for effect of test site ($F=1.98$, df 6,162, $p<0.071$), a significant effect of trial block (1 vs. 4: $F=430.68$, df 1,162, $p<0.0001$), and a significant interaction of site $\times$ block ($F=2.90$, df 6,162, $p=0.01$). Rather than a difference in habituation per se, this interaction reflected a significant difference in block 1 startle magnitude across sites ($F=2.73$, df 6,162, $p<0.015$), based on significant differences between site B and sites A ($p<0.0009$), E ($p<0.001$) and F ($p<0.05$). Consistent with this, ANOVA of *percent* startle habituation revealed no main effect of test site ($F=1.06$, df 6,162, ns).

Post-hoc comparisons examined the impact of inter-site sample or methodological differences on block 1 startle magnitude. Block 1 startle magnitude did not correlate significantly with either hearing threshold or electrode impedance (all r 's<0.1). Block 1 startle magnitude was significantly related to ethnicity; as described above for startle magnitude in blocks 2 and 3, this characteristic could account for site differences in block 1 startle magnitude. ANOVAs also detected no significant effects of sex, test order or split testing on block 1 startle magnitude or habituation, and no significant interactions, suggesting that these factors did not contribute significantly to site-related differences in these measures. Mean block 1 startle magnitude did not correlate significantly with ambient noise levels across sites ($r_s=0.37$, $p=0.38$).

3.3.3. Startle latency (Figure 5)—ANOVA of peak startle latency on pulse alone trials revealed no significant effect of test site ($F=1.74$, df 6,162, ns) or trial block ($F<1$), a significant effect of eye side (R > L: $F=6.46$, df 1,162, $p<0.015$), and no significant 2- or 3-way interactions. Site differences were also not evident for reflex onset latency ($F=1.04$, df 6,155, ns), and lateralized differences did not reach statistical significance ($F=2.45$, df 1,155, ns). Reflex latency normally becomes smaller (“latency facilitation” (Hoffman and Searle 1968; Ison et al. 1973)) when startling stimuli are preceded by short interval prepulses, and this latency facilitation was assessed by comparing peak latency across all trial types (PA and prepulse+pulse). Across all trial types, ANOVA of peak startle latency revealed no significant main effect of test site ($F=1.83$, df 6,144, ns) and significant main effects of sex (men > women: $F=6.31$, df 1,149, $p<0.015$) and trial type ($F=143.77$, df 3,144, $p<0.0001$), the latter reflecting the expected reduction in startle latency for 30, 60 and 120 ms prepulse conditions compared to PA trials (all p 's<0.0001). The effect of eye-side remained significant (R>L: $F=11.30$, df 1,144, $p=0.001$), and a significant interaction of site by eye side reflected substantial R>L asymmetry at some test sites (e.g. site C: mean (SEM) R vs. L (ms)=62.7 (0.80) vs. 59.9 (0.81); $p<0.002$) but not others (e.g. site B: mean (SEM) R vs. L (ms) = 57.8 (0.47) vs. 58.4 (0.50); ns). Similar patterns were observed for onset startle latency: no significant effect of site ($F=1.39$, df 1,137, ns), significant effects of sex (male > women: $F=6.31$, df 1,149, $p<0.015$) and trial type ($F=55.24$, df 3,411, $p<0.0001$, reflecting significant facilitation for 30 and 60 ms

prepulse trials (p 's<0.0001)) and side (R>L: $F=7.61$, df 1, 137, $p<0.007$), though the site by eye side interaction did not reach significance ($F=1.27$, df 6,137, ns).

3.3.4. PPI (Figure 6)—ANOVA of %PPI revealed no significant effect of test site ($F<1$), but significant effects of prepulse interval (120 ms > 60 ms > 30 ms; $F=68.02$, df 2,324, $p<0.0001$) and eye side (R > L; $F=7.78$, df 1,162, $p<0.006$), though not trial block ($F=1.83$, df 1,162, ns). There were no other significant 2-, 3- or 4-way interactions. Despite the lack of significant main effect of test site, there was a substantial range of mean %PPI values across sites (maximum (SEM) vs. minimum (SEM) averaged across prepulse intervals = 47.08 (1.83) vs. 35.20 (1.81)).

We examined factors that might have contributed to this range of PPI values across sites. Mean %PPI did not correlate significantly with either hearing threshold or electrode impedance (r 's<0.03), or startle magnitude during blocks 2-3 ($r<0.01$), suggesting that these variables did not contribute substantially to inter-site variability in PPI. Neither WRAT scores nor ethnicity were significantly related to PPI values. Across sites, mean PPI did not correlate significantly with ambient noise levels ($r_s = -0.03$, $p=0.96$). ANOVA across all sites revealed a significant effect of prepulse interval ($F=55.76$, df 2,322, $p<0.0001$), and a significant interaction of sex $\times$ test order ($F=5.70$, df 1,161, $p<0.02$; Figure 6B), but no significant effect of sex ($F<1$), test order ($F=2.00$, df 1,161, ns), or split testing ($F<1$), and no other significant 2-, 3- or 4-way interactions. Post-hoc comparisons revealed that the interaction of sex and test order reflected the expected pattern of significantly greater PPI in men than in women in test order A ($F=6.16$, df 1,99, $p<0.01$; Figure 6B), but an opposite trend in test order B ($F=2.28$, df 1,66, ns). Post-hoc comparisons revealed that, while PPI in women did not differ across test orders, PPI in men was significantly greater in test order A vs. B ($F=15.36$, df 1,70, $p<0.0003$). In considering possible factors accounting for this effect of test order on PPI in men, this sex-specific pattern did not appear to reflect the influence of smoking habits (smoking $\times$ test order: $F < 1$) or test site (site $\times$ test order: $F < 1$, for all sites except site C, which used only one test order).

4. Discussion

This study was an initial effort to evaluate the feasibility of multi-site measures of acoustic startle and startle plasticity in a community comparison sample, with an ultimate goal of using this multi-site approach to identify genes responsible for startle and PPI abnormalities in clinical populations. We assessed the consistency of startle and PPI measures in CCS across 7 geographically distributed testing sites, and examined differences in methodology and subject characteristics that might have contributed to the variance in these measures observed across sites. Substantial efforts were undertaken to ensure that subject recruitment, data acquisition and data analysis were standardized across the 7 sites. Each of the 7 sites in this study had many years of experience in subject recruitment and testing in psychophysiological measures, and one site (UCSD) had over 25 years of experience in measures of acoustic startle and PPI in normal and patient populations. Testers from all sites were trained during a several-day session, and testing instructions were distributed in the forms of manuals and videotapes (cf. Calkins et al. 2006). Equipment was assembled and calibrated at one site to enhance uniformity, and then distributed to the remaining 6 sites. Substantial efforts were dedicated towards quality assurance, and new personnel required "certification" before they were permitted to collect data. Annual re-training and re-certification procedures were also performed for all testers.

Adding to the standardization process, all waveforms were analyzed at a central site; while almost all waveform analyses are automated, all non-automated assessments (e.g. visual inspection of trials automatically marked as artifact) were made by a single investigator (JS), who remained blind to all subject characteristics. By utilizing multiple sites for data collection, it is possible to ascertain and test a large number of subjects from rare populations (e.g. large

schizophrenia pedigrees), and thereby overcome what would otherwise be a rate-limiting step in these types of studies. In contrast, data processing and analysis is not rate-limiting, and is most efficiently standardized at one site with expertise in a particular measure.

Through the QA process, a number of differences in experimental methodology were detected across sites. Some reflected methodologies previously used at a particular site, which were then applied to the present study (e.g. electrode type or placement, use of a video camera, split testing). Geographic or socioeconomic differences or differences in participant recruitment practices across sites may have contributed to differences in the sex and ethnic distributions, WRAT performance and prevalence of psychopathology in CCS across sites. Other differences likely reflected user-specific patterns (e.g. styles of skin abrasion and gel application can impact electrode impedance), equipment or test environment differences (e.g. hearing threshold), and others reflected user error (e.g. failure to balance test orders, accidental reversal of electrode channels). Among the methodological differences that could be quantified or categorized, none could independently account for the observed site variability in key dependent measures.

Given the complexities of this study, it is notable that no statistically significant differences were detected across the 7 sites in any of the four primary measures: startle magnitude (during blocks 2-3), reflex habituation, reflex latency / latency facilitation, and PPI. Significant differences across sites were detected only in block 1 startle magnitude. Nonetheless, a closer inspection reveals non-trivial variability across the 7 sites in each of these measures. For example, we examined %PPI at the 60 ms prepulse interval, where differences between medicated schizophrenia patients and CCS are often most robust (Braff et al. 1978, 2005; Leumann et al. 2002; Swerdlow et al. 2006; Weike et al. 2000). At this interval, the effect size (Cohen's *d*) for site differences in PPI in 8 out of the 21 pair-wise contrasts among the 7 sites in this study (A vs. B, A vs. C, A vs. D, etc.) exceeds that for the comparison between schizophrenia patients vs. CCS tested at one site (0.24) in a recent single-site study of 103 patients and 66 CCS (Swerdlow et al. 2006). Five of these pair-wise comparisons reached at least small effect sizes ($d > 0.3$), and one reached a medium effect size ($d > 0.5$). Despite this variability across sites, there was some evidence that the total sample variability of this study of %PPI for 60 ms intervals - as reflected by the standard deviation (37.82) - compares favorably with the variability of 60 ms PPI collected in 70 consecutive CCS at one test site over a comparable time interval (site A; $SD = 37.95$, unpublished observation). Furthermore, correcting alpha for multiple comparisons, no pair-wise contrasts among the 7 sites in this study reached statistical significance for this key measure (one pair reached $p < 0.05$, a second reached $p < 0.065$, and a third reached $p < 0.10$).

Post-hoc analyses narrowed the list of sources contributing to variability of measures across sites. Site-related patterns of startle variables do not appear to be related to differences in hearing threshold, electrode impedance, the use of split testing or levels of ambient room noise. Sex differences were evident in measures of both latency facilitation and PPI, the latter interacting with test order. Because both sex distributions and test order differed significantly across the 7 test sites, it is possible that these variables contributed significantly to the overall cross-site variability of latency facilitation and PPI.

In assessing the feasibility of multi-site measures of startle and PPI, we considered whether the patterns of findings from an integration of data from 7 test sites are consistent with those among the literature of single-site studies. This is clearly the case with many of the key measures: for example, the observed patterns of startle habituation, and the effects of prepulses and prepulse intervals on startle magnitude and latency are consistent with numerous previous reports (cf. Braff et al. 2001). Lateralized ($R > L$) patterns of PPI among CCS, though less often examined, have also been reported previously (e.g. Cadenhead et al. 2000).

Other findings from the present study do not agree perfectly with those previously reported in single-site studies. For example, sex differences have been reported previously in startle measures, particularly PPI (Jovanovic et al. 2004; Swerdlow et al. 1993, 1997). In the present study, however, the pattern usually reported in the literature - male > female PPI - was evident only in test order A, and not in test order B. Startle testing in order A most closely resembles that used in “stand alone” single-site startle studies. In contrast, PPI testing in order B was preceded by several hours of neurocognitive and other measures that might lead to participant fatigue. The substantial reduction in male PPI in order B vs. A is consistent with recent evidence linking fatigue with reduced PPI in men (van der Linden et al. 2006). No comparable published data addresses the effects of fatigue on PPI in women. It may also be relevant that we previously failed to detect male > female PPI levels when startle was measured in the context of a cognitive “challenge” (Talledo et al. 2005). Conceivably, whatever accounts for the fact that cognitive tasks mitigate male > female PPI differences (observed in uninstructed, automatic measures of PPI) might contribute to our failure to detect these sex differences in order B, where PPI testing follows several hours of neurocognitive batteries. While this study cannot elucidate the mechanisms underlying such an effect, the present findings at least suggest that test sequence should be controlled in multi-site PPI studies. Certainly, this would apply not only to studies with CCS, but also to those with patient populations.

Any multi-site startle study would benefit from this type of initial and continuing on-site standardization procedures used in the present study to detect a number of methodological differences across sites. However, with the increasing facility of web-based video monitoring, future studies might employ more “real time” supervision by a central QA test site. Even at sites with substantial experience in psychophysiological testing, methodological elements with long institutional histories can be introduced unknowingly into startle testing protocols, and create potential sources of cross-site variance. Presumably, this process might also occur in multi-site studies using other psychophysiological or neuropsychological measures. Conceivably, real-time monitoring might help detect this inadvertent contamination of methods in a manner that is more sensitive than the annual QA visits used here.

In summary, the present findings suggest that, even with substantial energy and detailed planning directed towards methodological standardization and quality assurance among an experienced group of investigators at different sites, startle data collected from community comparison subjects exhibits variability across sites. Based on our experience to date, a list of recommendations for future multi-site startle studies can be found in Table 4. Some sources of variability can be controlled relatively easily, including equal sex distributions among test groups and the use of a standardized testing order that includes startle testing relatively early in any test battery. Other sources of variability are more difficult to identify, but may include features of the testing environment or participant population. Some sources of variability might diminish as sites gain expertise through the testing of larger samples. However, in multi-site studies necessitated by the relative scarcity of a particular test population, test samples may remain small at any single site; infrequent testing and normal personnel turnover will complicate effective quality assurance and make it difficult to gain expertise in startle measures. Conceivably, given the amount of variability detected under the conditions of the present study, considerable thought in the design of startle studies should be given to the benefits of using many geographically dispersed test sites, each testing relatively smaller numbers of subjects, versus the benefits of using fewer test sites, each testing a larger number of participants and clinical subjects from larger geographic catchment areas.

Acknowledgements

We thank the following key personnel for their dedicated efforts to the COGS: Harvard University: Lynda Jacobs, Monica Landi, Erica Lee, Andrea Roe, Frances Schopick, and Alison Thomas. Mount Sinai School of Medicine: Rui Ferreira, Robert Fieo, Fran Schopick, Christopher Smith, Rebecca West. University of California Los Angeles:

William Horan, Mark Sergi. University of California San Diego: Andras Kovach, Katrin Meyer-Gomes, Barbara Haugeland, Kari Tweedale, Sheldrick Holmes, Emmeline Crowley. University of Colorado: Jamey Ellis, Jeff Hollis, Vicki Pender, Bernadette Sullivan, Bettye Clement, Christopher Cason, Alexis Ritvo. University of Pennsylvania: Alexandra Duncan Ramos, Jarrod Gutman, Carla Ann Henry, Paul Hughett, Jennifer Jackson, Adrienne Mishkin, J. Dan Ragland, Leslie Ramsey, David Rice, Jan Richard, Devon Seward, Felipe Silva and Robert Witalec. University of Washington: Kate B. Alvey, Andrew C. David, Sean P. Meichle, Denise O. Pritzl.

Role of the Funding Source. Supported by the National Institute of Mental Health (NIMH), via grants listed below by site. The NIMH had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. Funding sources for this work were: Harvard University: RO1-MH065562, MH43518, Commonwealth Research Center of the Massachusetts Department of Mental Health, Mount Sinai School of Medicine: RO1-MH065554; University of California Los Angeles: RO1-MH65707; University of California San Diego: RO1-MH065571, MH01436 (NRS); University of Colorado: RO1-MH65588; University of Pennsylvania: RO1-MH65578; University of Washington: RO1-MH65558.

References

- Andreasen, NC. The scale for the assessment of negative symptoms (SANS). University of Iowa; Iowa City, IA: 1984.
- Andreasen, NC. The scale for the assessment of positive symptoms (SAPS). University of Iowa; Iowa City: 1984.
- Anokhin AP, Heath AC, Myers E, Ralano A, Wood S. Genetic influences on prepulse inhibition of startle reflex in humans. *Neurosci. Lett* 2003;353:45–8. [PubMed: 14642434]
- Braff D, Stone C, Callaway E, Geyer M, Glick I, Bali L. Prestimulus effects on human startle reflex in normals and schizophrenics. *Psychophysiology* 1978;15:339–343. [PubMed: 693742]
- Braff, DL.; Freedman, R. *Neuropsychopharmacology: The Fifth Generation of Progress*. Lippincott Williams & Wilkins; Baltimore: 2002. The importance of endophenotypes in studies of the genetics of schizophrenia; p. 703-716.
- Braff D, Freedman R, Schork N, Gottesman I. Deconstructing schizophrenia: An overview of the use of endophenotypes in order to understand a complex disorder. *Schizophrenia Bulletin*. 2006in press
- Braff DL, Geyer MA, Light GA, Sprock J, Perry W, Cadenhead KC, Swerdlow NR. Impact of prepulse characteristics on the detection of sensorimotor gating deficits in schizophrenia. *Schizophr. Res* 2001a; 49:171–178. [PubMed: 11343875]
- Braff DL, Geyer MA, Swerdlow NR. Human studies of prepulse inhibition of startle: Normal subjects, patient groups, and pharmacological studies. *Psychopharmacology* 2001b;156:234–258. [PubMed: 11549226]
- Braff DL, Grillon C, Geyer M. Gating and habituation of the startle reflex in schizophrenic patients. *Arch. Gen. Psychiatry* 1992;49:206–215. [PubMed: 1567275]
- Braff DL, Light GA. The use of neurophysiological endophenotypes to understand the genetic basis of schizophrenia. *Dialogues Clin Neurosci* 2005;7:125–35. [PubMed: 16262208]
- Braff DL, Light GA, Ellwanger J, Sprock J, Swerdlow NR. Female schizophrenia patients have prepulse inhibition deficits. *Biol. Psychiatry* 2005;57:817–20. [PubMed: 15820241]
- Braff DL, Swerdlow NR, Geyer MA. Symptom correlates of prepulse inhibition deficits in male schizophrenic patients. *Am. J. Psychiatry* 1999;156:596–602. [PubMed: 10200740]
- Blumenthal TD, Cuthbert BN, Filion DL, Hackley S, Lipp OV, van Boxtel A. Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology* 2005;42:1–15. [PubMed: 15720576]
- Cadenhead KS, Carasso BS, 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;45:360–4. [PubMed: 10023514]
- Cadenhead KS, Swerdlow NR, Shafer KM, Diaz M, Braff DL. Modulation of the startle response and startle laterality in relatives of schizophrenia patients and schizotypal personality disordered subjects: Evidence of inhibitory deficits. *Am. J. Psychiatry* 2000;157:1660–1668. [PubMed: 11007721]
- Calkins ME, Dobie DJ, Cadenhead KS, Olincy A, Freedman R, Green MF, Greenwood TA, Gur RE, Light GA, Mintz J, Nuechterlein KH, Radant AD, Schork NJ, Seidman LJ, Siever LJ, Silverman JM, Stone WS, Swerdlow NR, Tsuang DW, Tsuang MT, Turetsky BI, Braff DL. The Consortium on the

- Genetics of Endophenotypes in Schizophrenia (COGS): “Model” recruitment, assessment, and endophenotyping methods for a multi-site collaboration. *Schiz Bull.* 2006in press
- Castellanos FX, Fine EJ, Kaysen DL, Kozuch PL, Hamburger SD, Rapoport JL, Hallett M. Sensorimotor gating in boys with Tourette's Syndrome and ADHD: Preliminary results. *Biol. Psychiatry* 1996;39:33–41. [PubMed: 8719124]
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. Second Edition. Lawrence Erlbaum; Hillsdale: 1988.
- Davis M. Sensitization of the rat startle response by noise. *J. Comp. Physiol. Psychol* 1974;87:571–81. [PubMed: 4415724]
- Francis DD, Szegda K, Campbell G, Martin WD, Insel TR. Epigenetic sources of behavioral differences in mice. *Nat. Neurosci* 2003;6:445–6. [PubMed: 12665797]
- Geyer MA, Braff DL. Habituation of the Blink reflex in normals and schizophrenic patients. *Psychophysiology* 1982;19:1–6. [PubMed: 7058230]
- Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am. J. Psychiatry* 2003;160:636–45. [PubMed: 12668349]
- Graham F. The more or less startling effects of weak prestimuli. *Psychophysiol* 1975;12:238–248.
- Gur RC, Ragland JD, Moberg PJ, Turner TH, Bilker WB, Kohler C, Siegel SJ, Gur RE. Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacol* 2001a;25:766–776.
- Gur RC, Ragland JD, Moberg PJ, Bilker WB, Kohler C, Siegel SJ, Gur RE. Computerized neurocognitive scanning: II. The profile of schizophrenia. *Neuropsychopharmacol* 2001b;25:777–788.
- Hall R. Global assessment of functioning: A modified scale. *Psychosomatics* 1995;36:267–275. [PubMed: 7638314]
- Hoffman HS, Flesher M. Startle reaction: Modification by background acoustic stimulation. *Science* 1963;141:928–30. [PubMed: 14043340]
- Hoffman HS, Searle JL. Acoustic and temporal factors in the evocation of startle. *J. Acoust. Soc. Am* 1968;43:269–282. [PubMed: 5636789]
- Ison JR, McAdam DW, Hammond GR. Latency and amplitude changes in the acoustic startle reflex produced by variations in auditory prestimulation. *Physiol. Behav* 1973;10:1035–1039. [PubMed: 4717640]
- Joover R, Zarate JM, Rouleau GA, Skamene E, Boksa P. Provisional mapping of quantitative trait loci modulating the acoustic startle response and prepulse inhibition of acoustic startle. *Neuropsychopharmacology* 2002;27:765–8. [PubMed: 12431851]
- Jovanovic T, Szilagyi S, Chakravorty S, Fiallos AM, Lewison BJ, Parwani A, Schwartz MP, Gonzenbach S, Rotrosen JP, Duncan EJ. Menstrual cycle phase effects on prepulse inhibition of acoustic startle. *Psychophysiology* 2004;41:401–6. [PubMed: 15102125]
- Kumari V, Das M, Zachariah E, Ettinger U, Sharma T. Reduced prepulse inhibition in unaffected siblings of schizophrenia patients. *Psychophysiology* 2005;42:588–94. [PubMed: 16176381]
- Kumari V, Gray JA, Geyer MA, ffytche D, Soni W, Mitterschiffthaler MT, Vythelingum GN, Simmons A, Williams SC, Sharma T. Neural correlates of tactile prepulse inhibition: a functional MRI study in normal and schizophrenic subjects. *Psychiatry Res* 2003;122:99–113. [PubMed: 12714174]
- Leumann L, Feldon J, Vollenweider FX, Ludewig K. Effects of typical and atypical antipsychotics on prepulse inhibition and latent inhibition in chronic schizophrenia. *Biol. Psychiatry* 2002;52:729–739. [PubMed: 12372664]
- Nurnberger JI Jr, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch. Gen. Psychiatry* 1994;51:849–59. [PubMed: 7944874]
- Postma P, Gray JA, Sharma T, Geyer M, Mehrotra R, Das M, Zachariah E, Hines M, Williams SC, Kumari V. A behavioural and functional neuroimaging investigation into the effects of nicotine on sensorimotor gating in healthy subjects and persons with schizophrenia. *Psychopharmacology* 2006;184:589–99. [PubMed: 16456657]
- Swerdlow NR, Braff DL, Geyer MA. Animal models of deficient sensorimotor gating: What we know, what we think we know, and what we hope to know soon. *Behav. Pharmacol* 2000;111:185–204. [PubMed: 11103873]

- Swerdlow NR, Geyer MA, Braff DL. Neural circuitry of prepulse inhibition of startle in the rat: Current knowledge and future challenges. *Psychopharmacology* 2001a;156:194–215. [PubMed: 11549223]
- Swerdlow NR, Hartman PL, Auerbach PP. Changes in sensorimotor inhibition across the menstrual cycle: Implications for neuropsychiatric disorders. *Biol. Psychiatry* 1997;41:452–460. [PubMed: 9034539]
- Swerdlow NR, Karban B, Ploum Y, Sharp R, Geyer MA, Eastvold A. Tactile prepuff inhibition of startle in children with Tourette's syndrome: in search of an “fMRI-friendly” startle paradigm. *Biol. Psychiatry* 2001b;50:578–85. [PubMed: 11690592]
- Swerdlow NR, Light GA, Cadenhead KC, Sprock J, Hsieh MH, Braff DL. Startle gating deficits in a large cohort of patients with schizophrenia: Relationship to medications, symptoms, neurocognition and level of function. *Arch. Gen. Psychiatry* 2006;63:1325–35. [PubMed: 17146007]
- Swerdlow NR, Monroe SM, Hartston HJ, Braff DL, Geyer MA, Auerbach PP. Men are more inhibited than women by weak prepulses. *Biol. Psychiatry* 1993;34:253–261. [PubMed: 8399822]
- Swerdlow NR, Shoemaker JM, Auerbach PP, Pitcher L, Goins J, Platten A. Heritable differences in the dopaminergic regulation of sensorimotor gating: II. Temporal, pharmacologic and generational analyses of apomorphine effects on prepulse inhibition. *Psychopharmacol* 2004;174:452–462.
- Swerdlow NR, Talledo JA, Braff DL. Startle modulation in Caucasian-Americans and Asian-Americans: A prelude to genetic/endophenotypic studies across the “Pacific Rim”. *Psychiatric Genetics* 2005;15:61–65. [PubMed: 15722959]
- Talledo JA, Sutherland AN, Nagy D, Swerdlow NR. Gating while rating: Effects of cross-modal matching on prepulse inhibition of startle. *Proc. Soc for Neuroscience Program No. 767.1.* 2004
- Turetsky B, Calkins M, Light G, Olincy A, Radant A, Swerdlow N. Neurophysiological Endophenotypes of Schizophrenia: The Viability of Selected Candidate Measures. *Schizophrenia Bulletin*. 2006in press
- van der Linden D, Massar SA, Schellekens AF, Ellenbroek BA, Verkes RJ. Disrupted sensorimotor gating due to mental fatigue: Preliminary evidence. *Int J Psychophysiol* 2006;62:168–174. [PubMed: 16730824]
- Weike AI, Bauer U, Hamm AO. Effective neuroleptic medication removes prepulse inhibition deficits in schizophrenia patients. *Biol. Psychiatry* 2000;47:61–70. [PubMed: 10650450]
- Wilkinson, G. WRAT3 Wide Range Achievement Test, Administration Manual 1993 edition. Wide Range, Inc; USA: 1993.

Figure 1.

Consortium on the Genetics of Schizophrenia (COGS) structure and interactions. Figure from Calkins et al. (2006), with permission of author. Data were uploaded to a central Data Core website, managed by the UCLA site. Files were accepted only if they met specific size and format criteria. Individual data files were downloaded by the QA site (UCSD), blind to diagnosis and other demographic variables.

Figure 2.

Schematic representation of startle trial types (top) and session design (below). The startle test session included 74 active and 18 blank “no-stim” trials (interspersed throughout the session), and lasted 23.5 min, beginning with a 5-min acclimation period with 70-dB(A) SPL white noise that continued throughout the session. Startle stimuli were 40-msec 115-dB(A) SPL noise bursts (near-instantaneous rise time, est. 1 ms). Prepulses were 20 ms noise bursts 15-dB above a 70-dB(A) SPL white noise background, with prepulse onset 30, 60 or 120 ms prior to pulse onset; using slightly more intense 16 dB prepulses with this startle system, prepulse-associated EMG activity is <0.5 % of startle stimulus-induced levels. Five startle stimuli were presented at the beginning (Block 1) and end of the session (Block 4) to assess habituation. In Blocks 2-3, pulse alone and each of the 3 prepulse trial-types was pseudo-randomly intermixed (9 trials per condition per blocks; inter-trial intervals 11-19 s (mean=15 s)). In 18 blank “no-stim” trials, data were recorded without stimulus presentation, to assess basal EMG activity. Subjects were excluded for mean block 2 startle magnitude <10 startle units; this threshold was established to identify the lowest 10% of a normal distribution of startle magnitude values from >200 adult CCS tested at UCSD, using the present stimulus parameters. Applied to the present sample, it identified the lowest 11.7%.

Figure 3.

Mean (SEM) values (A) and individual distributions (B) for startle magnitude during session blocks 2 and 3, when PPI is measured, at the 7 COGS test sites (A-G). 1 startle unit = 1.31 μ V. The focus of this study was the comparison of startle measures across test sites, and thus separate analyses of each startle variable at each site are not described. Nonetheless, of the primary startle measures that involve a within-subject comparison -- habituation (startle magnitude in block 1 vs. 4), prepulse facilitation of latency (reduction of onset and peak latency on prepulse trials compared to pulse alone trials), and PPI (reduction of startle magnitude on prepulse + pulse trials compared to pulse alone trials) -- all are highly significant at each individual site.

A. **B.** **Figure 4.**

Startle habituation, shown as mean (SEM) startle magnitude during blocks 1 and 4 (A), and as a percent score (B), across the 7 COGS sites. 1 startle unit = 1.31 μ V. * = $p < 0.05$, B vs. A, E and F

Figure 5.

Peak startle latency (ms (SEM)), on PA trials (0 ms prepulse intervals), and 30, 60 and 120 ms prepulse+pulse trials, across the 7 COGS sites. Inset shows distinct patterns of peak reflex latency modulation in women vs. men. Despite the lack of significant site $\times$ trial type interaction for peak latency, inspection of the data revealed some inter-site variability in the patterns of latency facilitation across the different prepulse intervals. For example, peak latency for 120 ms prepulse trials was significantly slower than for 30 and 60 ms trials for some sites (A,B,C,G) but not others (D,F), and fastest latencies were detected with 30 ms prepulse trials at some sites, and with 60 ms prepulse trials at others. Neither hearing threshold nor electrode impedance correlated significantly with reflex latency on PA or any prepulse trials (all $r_s < 0.13$). No significant relationships were detected between reflex latency and either WRAT scores or ethnicity. Ambient noise levels also did not correlate significantly with mean peak latency for each site on PA ($r_s = -0.03$, $p = 0.93$) or on any prepulse trials (r_s 's = -0.24 - (-0.01) ; p 's = 0.54 - 0.97). ANOVA across all sites revealed a significant interaction of gender $\times$ trial

type ($F=3.33$, $df\ 3,447$, $p<0.02$). As seen in the inset, this interaction is reflected in patterns of reflex latency across sites: sites testing predominantly men (e.g. site F) exhibited latency patterns characteristic of males in the present study (minimal difference in latency across prepulse intervals), and those testing predominantly women (e.g. site A) exhibited latency patterns characteristic of females in the present study (less facilitation with 120 vs. 60 ms prepulse intervals).

Figure 6.

A. Mean %PPI (SEM) with 30, 60 and 120 ms prepulse+pulse trials across the 7 COGS sites.

B. %PPI in men and women, shown in test orders A and B. * = $p < 0.0003$, significant effect of test order in men; # = $p < 0.015$, men > women, test order A, after significant interaction of sex $\times$ test order, $p < 0.02$.

Table 1

Overall sample characteristics of 170 Community Comparison Subjects and differences across sites

Sex (M : F) *	72 : 98
Mean age (range)	36.4 (18 - 64)
Handedness, self-reported (%R : %L : %ambidextrous)	89 : 10 : 1
Mean years education (SD)	15.4 (2.6)
Ethnicity * - Caucasian: African American: Hispanic: Asian: Native American	122: 21: 14: 8: 5
Non-smoker (%)	86.0
No current Axis I diagnosis (%) *	95.9
No lifetime Axis I diagnosis (%)	73.5
No current psychotropic medication (%)	92.9
Mean GAF (SD) ¹	82.6 (9.3)
Mean BPRS (SD) ²	1.68 (1.0)
Mean WRAT3 (SD) ^{3*}	51.3 (4.6)

Site (n)	% male ^a	% Caucasian ^b	% No current Axis I ^c	WRAT3 (mean (SD)) ^d
A (36)	33	67	100	52.2 (3.4)
B (26)	31	81	100	50.9 (4.9)
C (17)	47	71	93.75	49.2 (5.2)
D (29)	65	50	100	50.1 (4.6)
E (24)	25	75	95.8	53.0 (3.7)
F (24)	71	63	95.8	50.0 (6.1)
G (23)	35	96	82.6	52.9 (3.4)

* Significant differences across sites:

^a p < 0.01^b p < 0.05^c p < 0.05^d p < 0.05¹ GAF - Global Assessment of Functioning Scale (Hall 1995)² BPRS - Brief Psychiatric Rating Scale (range 1-7)³ WRAT3 - Wide Range Achievement Test raw score, Third Edition (Wilkinson 1993)

Table 2**COGS Test Battery: Orders A and B**

Test Order A	Test Order B
P50	CVLT
AS	DS-CPT
PPI	LNS
Break	CPT-IP
SANS / SAPS	PENN Neuropsychological Battery
PENN Neuropsychological Battery	SANS/SAPS
CPT - IP	Break
LNS	PPI
CVLT	AS
DS-CPT	P50

Abbreviations:

AS: Antisaccade measurements

P50: P50 Event-related potentials

SANS: Scale for the Assessment of Negative Symptoms

SAPS: Scale for the Assessment of Positive Symptoms

PENN: University of Pennsylvania

LNS Letter - Number Span test

CVLT California Verbal Learning Test

CPT: continuous performance task (DS = degraded stimulus version; IP = identical pairs version)

Table 3

Methodological / experimental characteristic differences across sites

Site	Test Order % A ¹	Split test % ²	Ambient noise (dB(A))	Hearing Threshold ³ (dB(A); mean (SD))	Electrode Impedance ⁴ (kΩ; mean (SD))
A	53.0	0.0	42	12.5 (6.6)	5.9 (2.4)
B	61.6	54.2	43	11.3 (5.4)	7.3 (1.6)
C	94.0	6.3	42	17.5 (3.8)	6.0 (1.6)
D	35.0	100.0	36	10.5 (5.3)	5.1 (2.7)
E	66.7	25.0	34	9.3 (3.1)	8.1 (1.5)
F	58.3	5.0	45	20.3 (4.7)	4.7 (1.9)
G	60.9	27.8	36	7.0 (5.0)	5.4 (4.3)

Examples of methodological variations detected across sites during QA process

1. active electrodes spaced too far apart or not directly over orbicularis oculi
2. ground electrode placement on forehead instead of behind ear, or at incorrect mastoid location
3. electrode channels reversed (right vs. left)
4. participant positioned to look upwards into a video camera or at a picture instead of straight ahead at a spot on the wall
5. use of alternative electrode or electrode cuff types
6. inaccurate stimulus calibration
7. modified test instructions, e.g. "We will get your reaction to the static."
8. problems with on-site QA: not checking electrode impedance or test-room gauss levels, not monitoring EMG waveforms for noise, variations in electrode care, etc.

¹ p < 0.025

² p < 0.001

³ p < 0.0001

⁴ p < 0.0001 (comparisons 1-2 by Chi-Square; 3-4 by ANOVA)

Table 4

Recommendations for minimizing cross-site variability in multi-site studies of the startle reflex in CCS*

1	Site selection:
	<ul style="list-style-type: none"> utilize test sites with significant past experience in psychophysiological measures, and at least one site with substantial expertise in acquisition and analysis of startle data for sites with existing or past startle protocols, identify and emphasize the differences between this new protocol vs. others used at that site (to prevent methods “blurring”) use historical data from sites to predict likely demographic patterns and potential differences in ascertainment / recruitment practices across sites
2	Standardization:
	<ul style="list-style-type: none"> acquire and calibrate identical equipment at a single site prior to distribution to test sites specify supplies (e.g. electrodes, gel, etc.) so that identical materials are used across sites yearly (re)training and (re)certification on test procedures instructional videos, procedure checklists and manual describing the equipment, software, electrode placement, scripts for interacting with participants and task administration standardized training of diagnosticians and regular review of inclusion / exclusion criteria
3	Test administration:
	<ul style="list-style-type: none"> if used as part of an extended test battery, perform startle testing early in the test day (and thereby avoid “split testing”) once test order is identified, it should be maintained constant within and across all test sites
4	Monitoring:
	<ul style="list-style-type: none"> yearly site-by-site inspection by designated startle QA expert centralized waveform review and scoring by designated expert regular conference calls and consultation with the central QA site regular remote observation of a test procedure via web-cam link regular review to assess site differences in CCS demographics, including sex distribution, ethnic representation, current and lifetime Axis I diagnoses, and psychotropic medications; in advance of site selection monitor differences in testing conditions: ambient noise levels, electrical noise, impedance, hearing thresholds

*
modified from Calkins et al. (2006)