Crime and physical activity measures from the Safe and Fit Environments Study (SAFE) Psychometric properties across age groups

Scott C. Roesch (sroesch@sdsu.edu)
San Diego State University Malcolm A Love Library https://orcid.org/0000-0001-6574-1916

Christina M. Patch
University of California San Diego

Caterina G. Roman
Temple University

Terry L. Conway
University of California San Diego

Ralph B. Taylor
Temple University

Brian E. Saelens
University of Washington

Marc A. Adams
Arizona State University

Kelli L. Cain
University of California San Diego

Loki Loki Natarajan
University of California San Diego

James F. Sallis
University of California San Diego

Methodology

Keywords: Crime, Physical Activity, Measurement, Psychometrics, Age, Fear of Crime

Posted Date: May 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-29595/v1

Version of Record: A version of this preprint was published at Preventive Medicine Reports on June 1st, 2021. See the published version at https://doi.org/10.1016/j.pmedr.2021.101381.

Crime and physical activity measures from the Safe and Fit Environments Study (SAFE): Psychometric properties across age groups

Scott C. Roesch
Department of Psychology, San Diego State University, San Diego, CA, USA

Christina M. Patch
University of California, San Diego (Department of Family Medicine and Public Health) and San Diego State University (Department of Public Health), Joint Doctoral Program in Public Health, San Diego, CA, USA

Caterina G. Roman
Department of Criminal Justice, Temple University, Philadelphia, PA, USA

Terry L. Conway
Department of Family Medicine and Public Health, University of California, San Diego, CA, USA

Ralph B. Taylor
Department of Criminal Justice, Temple University, Philadelphia, PA, USA.

Brian E. Saelens
Department of Pediatrics and Psychiatry & Behavioral Sciences, Center for Chld Health, Behavior, and Development, University of Washington & Seattle Children's Research Institute, Seattle, WA, USA

Marc A. Adams
College of Health Solutions, Arizona State University, Tempe, AZ, USA

Kelli L. Cain
Department of Family Medicine and Public Health, University of California, San Diego, CA, USA.

Loki Natarajan
Division of Biostatistics and Bioinformatics, Department of Family Medicine and Public Health, University of California San Diego, CA, USA

James F. Sallis
Department of Family Medicine and Public Health, University of California, San Diego, CA, USA.

Coresponding author: Scott C. Roesch (sroesch@sdsu.edu)

Abstract

Valid and reliable measures are needed to better understand the relationship between physical activity and crime. This paper provides a comprehensive psychometric evaluation of measures developed in the Safe and Fit Environments (SAFE) Study to assess a crime-PA conceptual framework. In addition to assessing the basic psychometric properties of each measure (e.g., variable distributions [item/scale level], internal consistency reliability), this study formally examined the measurement validity and invariance of measures across four age groups using confirmatory factor analysis. The sample (n=2,173) included 336 Adolescents (aged 12-17), 532 Young adults (aged 18-39), 838 Middle Age Adults, and 467 Older Adults (aged 66+). The psychometric evaluation of (sub)scales showed consistent factorial validity and internal consistency reliability across the majority of the measures and across the four age groups. Specifically, 14 of the 17 measures displayed statistically and practically significant factor loadings and internal consistency values in the overall sample and across the age groups. The pattern of correlations for each (sub)scale with other (sub)scales/indexes largely did not exhibit redundancy across measures. The findings expanded upon the test-retest reliability evaluation reported in Patch et al. (2019) and suggest these measures can be used in predictive models.

Keywords: Crime, Physical Activity, Measurement, Psychometrics, Age, Fear of Crime

Background

The lack of consistent support for widely held hypotheses about crime's effect on physical activity (1,2) calls for an examination of the conceptual models, measures, outcomes, and analyses. The Safe and Fit Environments Study (SAFE; see 3) sought to evaluate the relations among crime, fear of crime, responses to crime, environmental contexts, and physical activity by developing an interdisciplinary (i.e., criminology, psychology, public health) multilevel conceptual framework with valid and reliable measures to test hypotheses derived from the framework. Measures were developed for each construct of the framework, with the goal of generating items and subscales suitable for use across age groups (Adolescent [12-17], Young Adult [18-39], Middle Age Adult [40-65], Older Adult [66+]). Table 1 briefly describes the constructs. Multiple measures of each construct were developed to operationalize each conceptual domain. These broader conceptual domains included variables (measures) at the (a) Micro Level, variables that represent individual factors (cognitive assessment of crime/safety, emotional responses to crime, behavioral responses to crime/victimization/physical activity, and personal experiences/victimization); (b) Meso Level, variables that represent perceived social control and social cohesion in participants' communities; and (c) Macro Level, variables that represent crime prevention through reducing opportunities for crime in the neighborhood context.

A systematic process was used to develop and adapt survey items and scales to provide operational definitions of the constructs in the SAFE conceptual framework (3). The survey development process was guided by several principles: 1) draw from existing crime and physical activity measures when possible, 2) develop new items as needed based on focus group and expert input, 3) design items to apply across age, gender, and socio-demographic groups, and 4)

minimize respondent burden by limiting length of scales and maintaining a common response format. The survey development process began with a non-systematic literature review to identify existing measures in the fields of criminology, sociology, and physical activity that related to constructs in the framework. Over the course of a year, an interdisciplinary team of experts met weekly to review existing measures and adapt them as necessary. When measures did not exist or were deemed inadequate to represent a content domain, items were generated by the team. The experts discussed the language used for items, response formats, and psychometrics (when available). Some items were adopted from existing measures without material change, but most were altered somewhat in content, format, or both.

Test-retest reliability information for these measures was provided on a subsample of participants (N = 176) by Patch et al. (3). The majority of these measures achieved adequate reliability across age groups. Given the construction of "new" measures for the SAFE study, it is vital to conduct an integrated psychometric evaluation to confirm the construct validity of each measure before these measures can be used in predictive models. Of specific interest in the current analysis is the invariance of these measures (and by implication, constructs) across age groups. The use of large sample sizes for each age group provides an opportunity to validate the majority of the measures and establish construct validity, internal consistency reliability, and concurrent/discriminant validity across the lifespan. Only after it has been demonstrated that the constructs of interest are being measured similarly and independently in each age group can one attribute substantive differences/associations to substantive factors rather than noncomparability of the measurement instrument.

There are multiple aspects to validity, conceptually unified under construct validation (4). The structural aspect of validity refers to the correspondence between the scoring of the

measurement instrument and the hypothesized constructs and involves examination at the level of the items (also referred to as measurement validity). The extent of evidence required for demonstrating validity depends on the types of inferences one wishes to make from the obtained scores, but at a minimum, the structure of the measure should be established. This is referred to as factorial validity, and is traditionally established using factor analysis (exploratory, confirmatory). When using measurement instruments that have been developed for different age groups, an added concern is the extent to which the measurement instrument has the same structure across age groups. At the initial level of analysis, this is referred to as configural or pattern invariance (5,6). In the present study, factor loadings in each age group were tested to determine if they are invariant (equivalent). If invariance is observed, one can infer that a measurement instrument yields scores that can be interpreted in a similar fashion across different populations or groups, in this case different age groups.

In the current study, a comprehensive psychometric evaluation of the SAFE measures was undertaken. In addition to assessing the basic psychometric properties of each measure (e.g., variable distributions, internal consistency reliability), this study formally examined the factorial validity and invariance of measures using confirmatory factor analysis (CFA). Not all measures from the SAFE study were amenable to CFA. Of the 25 measures identified in Table 1, 17 were subjected to psychometric evaluation that specifically included CFA; these measures are referred to as (sub)scales throughout the manuscript. Eight measures could not be evaluated using CFA for a number of reasons, including operationalization by only two items for a given measure, use of a binary response scale, or the lack of an underlying continuum represented by the response scale. These measures are referred to as (summative) indexes through the manuscript.

Method

Participants

Recruitment. The survey was administered to study participants of the Safe and Fit Environments Study (SAFE Study) recruited from four metropolitan US regions: Baltimore counties Maryland/Washington DC; Seattle/King County, Washington; San Diego County, California; and Phoenix/Maricopa County, Arizona. Most participants in the SAFE Study were re-recruited from one of four previous studies conducted by the same research team: the Neighborhood Quality of Life Study (adults; 7), the Senior Neighborhood Quality of Life Study (older adults; 8), the Teen Environment and Neighborhood Study (adolescents; 9), and the Neighborhood Impact on Kids Study (children; 10, 11). Recruitment of new participants was conducted in the same regions, with oversampling from high-crime and low-socioeconomic status areas, as well as from a new region (WalkIT Arizona; 12).

In the present SAFE Study, the research team sampled to achieve a balance of participants from high- and low-crime neighborhoods and across four current age groups: 2,173 participants comprised the sample, with 336 participants in the Adolescent group (12-17), 532 participants in the Young Adult group (18-39), 838 participants in the Middle Age Adult group (40-65), and 467 participants in the Older Adult group (66+). Additional demographic information is provided in Table 2.

Measures

A brief description of each (sub)scale and index is provided in Table 1. The complete survey and more information about item adaptation from original sources is available online at https://drjimsallis.org/measures.html (see also 3).

Statistical Approach

The data for each measure in Table 1 were subjected to basic psychometric analysis. For the 17 measures for which CFA was used, descriptive statistics at the item level were evaluated to identify items with low variability and non-normal distributions. Subsequent to this, CFA models were tested for each measure to establish the best-fitting measurement model in the overall sample and stratified by age group. Traditionally, the likelihood ratio chi-square test has been reported but sparingly used to determine whether a model fits well. This test statistic has been identified as unsatisfactory for numerous reasons, including the heavy reliance of this statistic on sample size (13). While the use of alternative descriptive fit indices and the values used to determine overall model fit is contentious (14,15), three descriptive fit indexes have been generally recommended (16): (a) the Comparative Fit Index (CFI; 17), a relative index of model fit with values > .90 indicating acceptable model fit; (b) the root mean square error of approximation (RMSEA; 17), an absolute index of overall model fit with values less than .08 indicative of acceptable model fit, and (c) standardized root mean-square residual (SRMR; 16), an absolute index of overall model fit with values less than .08 indicative of acceptable model fit. However, given the exploratory nature of this psychometric analysis, these more liberal threshold values were used to indicate reasonably acceptable fit. All CFA models were estimated with the MPlus software version 8.1 (19) and used both Maximum Likelihood-Robust and Weighted Least Squares estimation procedures.

Finally, descriptive statistics at the (sub)scale/index level were examined for all 25 measures to identify (sub)scales/indexes with low variability items and non-normal distributions. Correlations among all measures were examined to determine if redundancy (i.e., correlations > .70) existed. These correlations were evaluated for the overall sample and for each age group.

Results

Descriptive statistics, confirmatory factor analyses, and correlations among measures are presented in Table 3-7 for the overall sample and by age group. Specifically, the tables present item-level descriptive statistics (Table 3), standardized factor loadings from the CFAs and internal consistency coefficients (Table 4), (sub)scale- and index-level descriptive statistics (Table 5), and correlations among all (sub)scales in the overall sample (Table 6) and by age group (Table 7). The results are presented below by conceptual level of assessment (Macro, Meso, Micro).

Macro-level: Neighborhood Context

Means and standard deviations for the Crime Prevention through Environmental Design (CPTED) measures all indicated reasonable normality and sufficient variation at the item level and across age groups (Table 3). The configural invariance model fit reasonably well for CPTED Surveillance (RMSEA = .114 to .156, SRMR=.057 to .078, CFI=.900 to .970). However, the standardized factor loadings for two of the items were not practically significant (loadings < .25) in the overall sample and for each age group ("There are many places in my neighborhood where criminals could wait for victims without being seen" and "The police patrol my neighborhood frequently").

Moreover, internal consistency values were relatively low (see Table 4). Conversely, all items for CPTED-Maintenance had large loadings, and internal consistency values were reasonable. At the subscale level, the descriptive statistics and normality for the four CPTED subscales exhibited similar means and standard deviations across age groups (see Table 5). No excessive non-normality was exhibited. Correlations with other study measures did not exhibit

redundancy with other measures and correlation patterns were similar across the four age groups (see Tables 6 and 7).

Meso-level: Social Dynamics

Means and standard deviations for the Collective Efficacy and Neighborhood Integration also indicated reasonable normality and sufficient variation at the item level. The configural invariance model fit reasonably well for both Collective Efficacy (RMSEA = .119 to .184, SRMR .067 to .081, CFI = .906 to .925) and Neighborhood Integration (RMSEA = .156 to .214, SRMR .042 to .067, CFI = .935 to .981). The standardized factor loadings for both measures were all large, significant, and similar across age groups. The internal consistency values were all large. At the scale level, the means and standard deviations were similar across age groups and non-normality was not evident. While the two measures within this domain did exhibit a strong correlation with each other (*r*s ranged from .54 to .57), correlations with other study measures did not exhibit redundancy with other measures, and correlation patterns were similar across the four age groups.

Micro-level: Individual Factors (Personal Experiences)

Means and standard deviations at the item level for the four Victimization subscales all indicated non-normality with significant positive skew. There was infrequent endorsement of victimization. The configural invariance model fit reasonably well for all four Victimization subscales: Recent (RMSEA = .001 to .069, SRMR = .009 to .159, CFI = .953 to 1), Ever (RMSEA = .008 to .110, SRMR = .015 to .073, CFI = .957 to 1), Witnessing Crime (RMSEA = .001 to .072, SRMR = .019 to .078, CFI = .971 to 1), Hearing about Crime (RMSEA = <.001 to .001, SRMR = .001 to .027, CFI = .999 to 1). The overwhelming majority of factor loadings were large, significant, and similar across age groups. The exception was a single item in the 12-

17 age group for the Recent subscale (item: In the past 12 months, how many times have you been the victim of property crimes [including theft, motor vehicle theft, burglary, vandalism]). This item did not significantly load (standardized value = .08). Internal consistency values were relatively low for the Recent and Past subscales but were consistently reasonable for the Witnessing Crime and Hearing about Crime subscales. At the subscale level, these measures continued to exhibit severe non-normality with significant positive skew.

The Crime Information Resources summated index exhibited an approximately normal distribution with significant variation. All measures within this domain exhibited similar means and standard deviations across age groups. Some correlations among the four Victimization subscales were strong, but none exceeded .65. Correlations with Crime Information Resources and other measures did not indicate redundancy.

Micro-level: Individual Factors (Cognitive Assessment of Crime)

Item-level means and standard deviations for Evaluation of Risks and Values/Incivilities indicated relatively low endorsement of some items; this did not, however, result in significant non-normality. The Street Efficacy measure approximated normality well. There was significant variation for all three measures. The configural invariance model fit reasonably well for Evaluation of Risks (RMSEA = .083 to .094, SRMR .045 to .056, CFI = .968 to .986), Values/Incivilities, (RMSEA = .040 to .060, SRMR .044 to .064, CFI = .962 to .977), and Street Efficacy (RMSEA = .041 to .141, SRMR .009 to .025, CFI = .988 to .999). The standardized factor loadings for all three measures were large, significant, and similar across age groups. The internal consistency values were also large. Similar to what was observed at the item-level, at the scale level, the means and standard deviations were similar across age groups, and non-normality was not evident. Significant and strong correlations were found between Evaluation of Risks and

Values/Incivilities (*r*s ranged from .47 to .65). Strong, positive relationships were also found between these two measures and No Behavioral Response (*r*s ranged from .48 to .64), but not strong enough to indicate redundancy. Correlations with other study measures did not exhibit redundancy with other measures, and correlation patterns were similar across the four age groups. No strong correlations were found for the Street Efficacy measure with other study measures, and the pattern of correlations was similar across age groups.

Micro-level: Individual Factors (Emotional Responses to Crime)

Means and standard deviations for the Fear of Crime measure indicated reasonable normality and sufficient variation at the item level. The configural invariance model fit reasonably well (RMSEA = .115 to .124, SRMR .056 to .070, CFI = .948 to .966). The standardized factor loadings for both measures were all large, significant, and similar across age groups. The internal consistency values were all large. At the scale level, the means and standard deviations were similar across age groups, and only mild non-normality was evident. This measure did exhibit a strong correlation with the Values/Incivilities measure (r = .65), but correlations with other study measures did not exhibit redundancy.

Micro-level: Individual Factors (Behavioral Responses to Crime)

Means and standard deviations for the Protective Behaviors measure indicated reasonable normality and sufficient variation at the item level. The configural invariance model fit reasonably well (RMSEA = .068 to .078, SRMR .050 to .080, CFI = .962 to .980). The standardized factor loadings for this measure were generally large, all were significant, and they were similar across age groups. The internal consistency values were all large. At the scale level, the means and standard deviations were similar across age groups but mild non-normality was evident across age groups. This measure exhibited strong correlations with Avoidant Behaviors

(Dark, Alone subscale and the Positive Avoidant Behaviors measure, respectively [*r*s ranged from .61 to .68]), but correlations with other study measures did not indicate redundancy.

Means and standard deviations at the item level for the four Avoidant Behaviors subscales all indicated some non-normality with positive skew. There was generally a low endorsement for the items from these subscales. In general, mild non-normality was evident for the majority of items. There was, however, significant non-normality (positive skew) for both the Daylight, Alone and Daylight, Others subscales in both the 40-65 and 66+ age groups; also for the Daylight, Others subscale in the 18-39 age group. The configural invariance model fit reasonably well for all four subscales: Daylight, Alone (RMSEA = .028 to .070, SRMR = .012 to .018, CFI = .996 to 1), Daylight, Others (RMSEA = .001 to .060, SRMR = .009 to .018, CFI = .995 to 1), Dark, Alone (RMSEA = .059 to .106, SRMR = .006 to .021, CFI = .995 to 1), Dark, Others (RMSEA = .024 to .082, SRMR = .010 to .021, CFI = .996 to 1). All standardized factor loadings were large, significant, and similar across age groups. Internal consistency values were also strong across subscale and age group. At the subscale level, these subscales exhibited mild non-normality, but not overly-severe. Correlations among the four subscales were highly redundant across age groups (*r*s ranged from .72 to .91). These subscales were not, however, redundant with other study measures.

For the summated indexes in this variable domain, both the News-Related Avoidant Behaviors index and the Obligatory Behaviors index were severely non-normal, with positive skew, in the two older age groups. Severe non-normality was also evident for all four age groups for the Community Participation index. The Positive Avoidant Behaviors summated index and the No Behavioral Response summated index indicated reasonable normality and variation at the scale level. While correlations between these summated indexes were similar across age groups,

some strong correlations were found between Positive Avoidant Behaviors and No Behavioral Response (*r*s ranged from .51 to .64) and Effect of Safety on Physical Activity measure and No Behavioral Response (*r*s = .63).

Discussion

The psychometric evaluation of the SAFE (sub)scales showed consistent factorial validity and internal consistency reliability across the majority of the measures, and importantly, across the four age groups. Therefore, the newly-developed and adapted single set of questions accurately measured target constructs for all groups, and the common measures can be used to facilitate analysis and interpretation of age-related patterns. This approach to simultaneous measurement development for the lifespan is the best way to determine whether common measures are feasible, and it is more time- and cost-efficient than a sequential development approach. Specifically, 14 of the 17 measures subjected to a test of factorial validity displayed statistically significant and strong factor loadings and internal consistency in the overall sample and across the age groups. The pattern of correlations for each (sub)scale with other (sub)scales and the 8 indexes did not exhibit any redundancy, with the exception of the Avoidant Behavior subscales. That is, each measure identified a degree of unique measurement variance in operationalizing a given construct.

While this psychometric evaluation was largely supportive of the validity and reliability of the SAFE measures, some items, and by implication measures, had weaknesses. The primary Macrolevel measure of neighborhood context, CPTED, could use improvement through additional item development. It should be emphasized that no existing scale or measures operationally defining the CPTED construct previously existed. The measures (four subscales) were developed to fully define physical design features (20). The CPTED-Surveillance subscale

in particular had lower factor loadings for two of the five items, and this was found across the four age groups. While the inclusion of these two items attenuated reliability to a degree, the overall reliability value is still fair given the number of items comprising the scale. Moreover, these two items are core concepts for CPTED construct. Given the conceptual strength of these items, and the minimal impact on reliability, these items can still be used but with great care. The Access Control and Territorial Reinforcement subscales, respectively, had limited psychometric evaluation. Given that these subscales were operationalized by only two items, factorial validity and internal consistency reliability could not be meaningfully assessed. The CPTED-Maintenance subscale, however, showed strong psychometric characteristics across the age groups.

Similar to the CPTED measure, the four Victimization subscales showed varying degrees of psychometric viability. All four subscales did exhibit strong factor validity. However, two of the four subscales, Recent and Past, respectively, exhibited a low degree of internal consistency. This was particularly pronounced for the Recent subscale. Not surprisingly, this/these measures also exhibited the most severe non-normality of all the measures. The low internal consistency is likely reflective of the low rates observed for some items from this/these specific subscales, and by implication the construct(s) being measured. For example, in the Adolescent age group, the item “In the past 12 months, how many times have you been the victim of property crimes [including theft, motor vehicle theft, burglary, vandalism]” from the Recent subscale did not significantly load on the factor in the CFA. This item is likely not as relevant for this younger age group as it is for the 3 older age groups. The low internal consistency values could also be a function of how Cronbach’s alpha is calculated. In addition to the inter-item correlations, Cronbach’s alpha is also a function of the number of items that compose a scale. In this case,

there are 4 items per subscale. The magnitude of the standardized factor loadings, except for the Adolescent age group, all reflect fairly strong inter-item correlations. However, when the number of items is factored in to the calculation of alpha, the overall internal consistency value is necessarily attenuated. Given this, applied researchers should use these subscales with care.

The four Avoidant Behaviors subscales (Daylight, Alone; Daylight, Others; Dark, Alone; Dark, Others) all exhibited factorial validity and reliability across the age groups. However, the correlations among the four subscales indicated strong redundancy. For the majority of the correlations these values exceeded .70. This indicates a high degree of collinearity that could impact predictive models in which all subscales are added as simultaneous predictor variables. For this/these measures, participants' responses did not reflect a difference between whether or not the target avoidant behavior was applicable to themselves/others or daylight/dark. Interestingly, these subscales were not redundant with other study measures, including related Avoidant measures from the same variable domain (e.g., Positive Avoidant Behaviors). At this point, it would not be prudent to use these four subscales in similar analyses. However, it does appear that the items from these four subscales could be aggregated for use as a General Avoidant scale.

This study is not without other general weaknesses. Not all of the SAFE measures could be validated using a factor-analytic approach. Many of the indexes were measured using a dichotomous rating scale. First, factor analysis is based on the assumption that items have an underlying continuum, even if measured in a binary format. This assumption could not be met with some of these indexes. Second, some of the indexes (e.g., Crime Resources Index) are more appropriately conceptualized and operationally defined using a checklist format, and thus are not represented meaningfully as a factor (or latent variable). Third, some measures were

operationalized using as few as two items. These measures are similarly not amenable to factor-analytic evaluation to establish construct validity or reliability. However, these measures all showed significant unique measurement variance as reflected by the relatively small correlations found with the sub(scales) and other summated indexes. While these indexes were not amenable to formal factor-analytic procedures, their measurement form and properties suggest that they are reasonable measures to use in future studies. Finally, several sub(scales) and summated indices exhibited non-normality in the form of positive skewness at both the item- and scale level. This should not preclude the use of these measures in future studies. Measures of victimization, for example, are inherently skewed given their infrequent nature, but are vitally important in conceptual models of crime and physical activity (3).

Conclusions

In conclusion, these analyses supported the psychometric characteristics for the measures from SAFE, with the exceptions identified above. Importantly, the measurement properties were shown to be invariant (equal) across the four age groups of interest, thus facilitating cross-group comparisons in predictive models. This conclusion is consistent with the preliminary test-retest reliability evaluation reported in Patch et al. (3). Although not reported due to space limitations, the findings reported herein were confirmed using additional statistical procedures specific to categorical data analysis (including item response theory) and accounting for clustering at the block group level. While we recommend these measures for use (with the caveat of not simultaneously among the Avoidant Behaviors measures), continued psychometric evaluation, and implementation of other forms of measurement invariance (21), should be pursued to further refine these instruments and enhance their validity and reliability. The detailed evaluation of the

measures, especially identification of weaknesses, will inform interpretation of results in subsequent papers.

References

1. Carver A, Timperio A, Crawford D. Playing it safe: The influence of neighbourhood safety on children's physical activity—A review. Hlth Place.2008;14(2):217-227. doi:10.1016/j.healthplace.2007.06.00

2. Foster S, Giles-Corti B. The built environment, neighborhood crime and constrained physical activity: an exploration of inconsistent findings. Prev Med. 2008;47(3):241-251. doi: 10.1016/j.ypmed.2008.03.017

3. Patch CM, Roman CG, Conway TL, Taylor RB, Gavand KA, Saelens BE, Adams MA, Cain KL, Engelberg JK, Mayes L, Roesch SC, Sallis JM. Crime and physical activity: Development of a conceptual framework and measures. J Phys Act Hlth. 2019; 16(10):818-829. doi: 10.1123/jpah.2018-0405

4. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Amer Psy.1995;50(9):741–749. doi:10.1002/j.2333-8504.1994.tb01618.x

5. Roesch SC, Norman, GJ, Merz, EL, Sallis JF, Patrick K. Longitudinal measurement invariance of psychosocial measures in physical activity research: an application to adolescent data. J App Soc Psych. 2013; 43(4): 721-729. doi: 10.1111/jasp.12000

6. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organization research. Organizational Res Methods. 2000;3(1):4-70. doi:/10.1177/109442810031002

7. Sallis JF, Saelens BE, Frank LD, Conway TL, Slymen DJ, Cain KL, Chapman JE, Kerr. Neighborhood built environment and income: examining multiple health outcomes. Soc Sci Med. 2009;68(7):1285-1293. doi: 10.1016/j.socscimed.2009.01.017

8. King AC, Sallis JF, Frank LD, Saelens BE, Cain K, Conway TL, Chapman JE, Ahn DK, Kerr J. Aging in neighborhoods differing in walkability and income: associations with physical activity and obesity in older adults. Soc Sci Med. 2011;73(10) 1525-1533. doi: 10.1016/j.socscimed.2011.08.032

9. Carlson JA, Saelens BE, Kerr J, Schipperijn J, Conway TL, Frank LD, Chapman JE, Glanz K, Cain, KL, Sallis JF. Association between neighborhood walkability and GPS-measured walking, bicycling and vehicle time in adolescents. Health & Place. 2015;32:1-7. doi: 10.1016/j.healthplace.2014.12.008

10. Frank LD, Saelens BE, Chapman J, Sallis JF, Kerr J, Glanz K., ... Cain KL. Objective assessment of obesogenic environments in youth: geographic information system methods and spatial findings from the neighborhood impact on kids study. Amer J Prev Med. 2012; 42(5): e47-e55. doi: 10.1016/j.amepre.2012.02.006

11. Saelens BE, Sallis JF, Frank LD, Couch SC, Zhou C, Colburn T, Cain KL, Chapman J, Glanz K. Obesogenic neighborhood environments, child and parent obesity: the Neighborhood Impact on Kids study. Amer J Prev Med. 2012;42(5):e57-e64. doi: 10.1016/j.amepre.2012.02.008

12 Adams MA, Hurley JC, Phillips CB, Todd M, Angadia SS, Berardic V, Hovell MF, Hooke S. Rationale, design, and baseline characteristics of WalkITArizona: A factorial randomized trial testing adaptive goals and financial reinforcement to increase walking across higher and lower walkable neighborhoods. Contemporary Clinical Trials, 2019; 81(5);87-101. doi: 10.1016/j.cct.2019.05.001

13. Hoyle RH. Confirmatory factor analysis. In: Tinsley HEA, Brown SD, editors. Handbook of applied multivariate statistics and mathematical modeling. San Diego: Academic Press; 200. p. 465-497.

14. Bentler PM. On tests and indices for evaluating structural models. Pers Ind Dif. 2007;42(5):825-829. doi: 10.1016/j.paid.2006.09.024

15. Marsh HW, Hau KT, Wen Z. In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler's (1999) findings. Struct Equ Mod. 2004;22(3): 320–341. doi: 10.1207/s15328007sem1103_2

16. Bentler PM. Comparative fit indexes in structural models. Psych Bull. 1990;107(2):238-246. doi: 10.1037/0033-2909.107.2.238

17. Steiger JS. Structural model evaluation and modification: An interval estimation approach. Multivariate Beh Res. 1990;25(2):173-180. doi:10.1207/s15327906mbr2502_4

18. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Mod.1999;6(1):1-55. doi: 10.1080/10705519909540118

19. Muthén L. Muthén B. Mplus User's Guide (4th edition). Los Angeles: Muthén & Muthén; 2019.

20. Taylor, R. B. (2017). How do we get to causal clarity on physical environment-crime dynamics? In G. Bruinsma & S. Johnson (Eds.), *Oxford Handbook of Environmental Criminology*. New York: Oxford University Press.

21. Ployhart RE, Oswald, FL. Applications of mean and covariance structure analysis: Integrating correlational and experimental approaches. Organizational Research Methods; 2004;7(1):27–65. doi: /10.1177/1094428103259554

Abbreviations

SAFE: Safe and Fit Environments study

CFA: Confirmatory Factor Analysis

CFI: Comparative Fit Index

RMSEA: Root Mean Square Error of Approximation

SRMR: Standardized Root Mean-Square Residual

CPTED: Crime Prevention through Environmental Design

Declarations

Ethics approval and consent to participate

This study was reviewed and approved by the Institutional Review Board at the University of California San Diego under protocol #141227.

Consent for publication

Not applicable

Availability of data and materials: Data are housed at the University of California, San Diego. Data can be made available on request to Dr. James F. Sallis

Competing interests

The authors declare that they have no competing interests.

Funding

This study was funded by NHLBI grant #5R01HL117884.

Authors' Contributions

SCR was the primary author, contributed to the design of the study, designed and conducted analyses, compiled edits, and created the final version for submission. CP was co-lead in developing the measures, drafted sections of the paper and reviewed all drafts. CR was co-lead in developing the measures, contributed to design and implementation of the study, drafted sections of the paper, and reviewed all drafts. TLC contributed to design of the study, helped develop the measures, created and managed the database, contributed to design of measures and analyses, and reviewed all drafts. KC contributed to design of the study, contributed to design of measures, helped create the database, and reviewed all drafts. MAA contributed to the design of the study, helped develop the measures, and reviewed drafts of the paper. BES contributed to the design of the study, contributed to design of measures and analyses, and reviewed drafts of the paper. RT contributed to the design of the study, contributed to development of the measures, and reviewed drafts of the paper. LK contributed to the design of the study, contributed to planning analyses, and reviewed drafts of the paper. JFS conceptualized the study obtained funding, contributed to development of measures, and reviewed all drafts of the paper. All authors reviewed and approved this version.

Acknowledgements

Not applicable

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RoeschFINALSAFEmeasuresmenttables051720.pdf