HEALTHY KIDS COLORADO SURVEY

Technical Documentation

2021

Prepared by colorado school of public health

Contents

Overview of the 2021 Administration

The current report describes methods, including recruitment, administration, data cleaning and weighting for the 2021 Healthy Kids Colorado Survey (HKCS).

The 2021 administration was similar to previous administrations with a few differences worth highlighting. Following a supplemental administration in the fall of 2020, this was the first *statewide* administration of the HKCS since the onset of the COVID-19 pandemic and its impact on schools. For information on sampling, in 2021, the Centers for Disease Control and Prevention (CDC) drew a Youth Risk Behavior Survey (YBRS) sample of 40 public high schools, as was done in previous years. The Colorado School of Public Health (CSPH) Survey Team drew the regional/state high school sample of 146 schools, representing a smaller sample size than in 2019 as fewer high schools per region were selected (up to 7 per region, where possible). The CDC continued to provide the 1801/Healthy Schools high school sample and CSPH Survey Team sampled the middle school state sample and identified schools with a high concentration of Ute tribal membership for the high school sample (referred to as the Ute sample or Native American/American Indian oversample). See Table 2 for more details. Following the CDC's recommendation from prior years, alternative schools were not included in the sampling frame for the YRBS or HKCS samples.

Questionnaire Development and Design

The HKCS collects self-reported information from students in public middle and high schools. In 2021, the HKCS administered two online versions of the high school questionnaire (High School Module A, High School Module B), and one online middle school questionnaire. The two high school questionnaires shared a common core of questions, and each has different additional questions. This design enabled the HKCS to obtain information about more topics while keeping questionnaire length manageable for students. High school questionnaires were administered to students in grades 9-12, and middle school questionnaires were administered to middle and junior high school students in grades 6-8. English and Spanish versions were available for each questionnaire. School districts could make local modifications, either adding or removing questions, with some constraints. Results for any added questions were provided back to the respective schools and districts. The YRBS was administered separately from the HKCS, using paper surveys available in both English and Spanish, whereas the HKCS was administered exclusively online.

Changes to the 2021 questionnaires were solicited by outreach to a variety of partners and stakeholders, especially representing public health and education. Individuals or organizations were invited to submit questions for addition, modification, or removal. Members of the School and Youth Survey Steering Committee were asked to assess the value of new or modified questions and the potential impact of any question removal. The Steering Committee approved all final changes to the HKCS questionnaires.

- **HKCS High School Module A:** 88 core questions plus 32 module-specific questions; 120 total questions
- **HKCS High School Module B:** 88 core questions plus 32 module-specific questions; 120 total questions
- **YRBS High School:** 87 total questions
- **HKCS Middle School:** 100 total questions

Sampling and Eligibility

HKCS Sampling

The HKCS used a two-stage clustered sampling design, and at the high school level, a stratified design was also employed. Schools sampled in the first stage (stratified by region, at the high school level) and classrooms sampled within selected schools in the second stage. For the first stage, the sampling frame was obtained from the Colorado Department of Education and listed all public high schools and middle schools, including charter schools. Alternative, special education, online, and home schools were excluded as ineligible. The high school frame contained schools that served students in grades 9-12. The middle school frame contained schools that served students in grades 6-8 (middle schools or junior high; elementary schools with a 6th grade were excluded). Schools serving students in both middle and high school grades were included in both middle and high school frames. The high school sampling frame was provided to the CDC and Westat, which selected the YRBS sample and the 1801/Healthy Schools samples. The CSPH Survey Team selected the stratified high school sample, the middle school sample, and the Ute tribal oversample.

For the second stage, the classroom sampling frame included each school's classrooms in a required subject or all classrooms during a particular period of the school day. Classrooms within each school were sorted by grade, then alphabetically by teacher last name, and classrooms with fewer than 5 students were combined with an adjacent class on the class list.

The YRBS and 1801 classroom samples were selected using random numbers provided by CDC. In the HKCS regional/state high school sample and middle school sample, schools were offered the choice between a census (all students participating) and a selection (only sampled classrooms participating). Generally, schools were encouraged to utilize a census when possible. For schools that elected to use a selection, we calculated average class size to determine how many classrooms would result in approximately 100 students per grade for high schools, and 50 students per grade for middle schools. For active consent schools, or by request, we oversampled by 50% in order to reduce the number of estimates that would be suppressed. For all samples, ineligible classes included special education (where students were unable to take the survey independently, given the anonymous nature of the survey), online classrooms (for YRBS only), and classrooms where all or most students did not speak English or Spanish.

Subsample and Supplemental Sample

A supplemental **Ute Tribal concentration oversample** included three high schools and three middle schools attended by large numbers of students belonging to the Ute Mountain Ute or the Southern Ute tribes. These schools were eligible to be included in the state sample. A supplemental **1801/Healthy Schools sample** included 7 high schools that receive federal funding for nutrition, physical activity, and the management of chronic conditions in schools. Some of these schools were also included in the state sample.

Administrative Approval and School Recruitment

Survey approval and recruitment occurred at multiple levels: state, district, school, parent/guardian, and student. At the state level, approval was attained from the Steering Committee, the Colorado Multiple Institutional Review Board, and the Educational Data Advisory Committee.

After state approvals were completed in the spring of 2021, each superintendent whose district included a sampled school received a letter and follow up communications via email or phone call requesting the district's participation. After a district agreed to participate, each principal of a sampled school was contacted to request the school's participation, explain the HKCS administration, and offer a monetary stipend ($300) and a school-specific report of results. Contact procedures were tailored when necessary to coordinate with other survey activities or district communications preferences.

Parental consent was obtained via letters sent to parents/guardians at least two weeks before survey administration, following standard school communication with parents. Teachers administering the survey showed the instructional video or read the written instructions aloud to students, and the same instructions were presented in writing at the beginning of the survey informing students that their participation was voluntary and anonymous. No personally identifiable information was collected. Table 1 summarizes the approval and consent processes.

Table 1: Approvals and consent summary, 2021 HKCS

State	• HKCS Steering Committee • Colorado Multiple Institutional Review Board • Educational Data Advisory Committee (EDAC)
School District	• Superintendent approval • District application approval, where required • Determination of active or passive parental consent
School	• Principal approval • Designation of survey coordinator • Support of survey coordinator to plan administration
Parent/Guardian	• Active or passive consent for student participation
Student	• Teacher instructs student that participation is voluntary • Teacher instructs student that the survey is anonymous • Student decides to participate in the survey

Data Collection and Processing

Data were collected from September 2021 through January 2022. Principals designated building-level survey coordinators to manage the online survey administration. Survey coordinators were electronically provided instructions on sampling classrooms (when applicable), obtaining parent/guardian consent, and tracking survey progress. An instructional video and script was provided to survey coordinators to play or read on the day of survey administration. Survey coordinators had access to registration forms, email communications, and the HKCS public website for all instructions.

Schools administering the HKCS did so online using Qualtrics. Each participating school was given a unique survey URL to provide to all students. The unique URL allowed for tracking school-specific data while maintaining student survey anonymity. Response rates were tracked throughout each schools' survey date or window of administration dates, and the survey coordinators were notified when their school was below a 60% student response rate in an effort to address non-response by following up with missing classroom data.

The 2021 Qualtrics survey utilized survey skip and exclusive logic to reduce exposure to non-relevant questions. Skip logic was added to eight questions in both high school modules A and B, and seven middle school questions. Up to 23 questions in the high school module A survey, 16 questions in the high school module B survey, and 14 question in the middle school survey could have been skipped. Exclusive logic was applied to select-all-that-apply questions to prevent students from selecting conflicting survey response options. Exclusive logic was applied to 11 high school module A questions, six high school module B questions, and four middle school questions.

For the YRBS classrooms that surveyed on paper, survey packets and answer sheets were shipped to schools. Once returned, materials were visually inspected for stray marks, incomplete erasures, or incomplete bubbling on answer sheets. The answer sheets were then shipped directly to Westat, on behalf of the CDC, for data processing.

For the HKCS, raw data were inspected to remove duplicate entries, edit out-of-scope values, and verify variable attributes (length, type, order). Student responses were "subverted" or set to missing if they had 20 or fewer total responses for the entire survey. Logical consistency was verified in accordance with the YRBS Data User's Guide. Additional logical consistency verifications were added for questions outside the core YRBS questions. Questions were checked against related questions to assess consistency, and invalid responses were set to missing values.

Weighting

Overview

The HKCS and other sample-based surveys use weights to account for unequal chances of completing the survey. Use of the weights in analysis allows the sample to support inferences about the whole population. The weights adjust for unequal sampling rates, school and student nonparticipation and nonresponse, and discrepancies between the overall sample and the population in grade, sex, and race/ethnicity.

Weights were constructed as a series of factors and final weights are products of the factors. High school data have three regional weights for estimating state and regional results: core questions (FWTRGN) and for questions on only Module A (FWTRGNA) or Module B

(FWTRGNB). Middle school data include one final weight for estimating statewide results. The "Guide to Data Analysis" section describes appropriate use of weights.

School Selection Probability and Weight (W_1)

The *sampling fraction* is the probability that the school would be selected (or sampled) by the CSPH Survey Team. The sampling fraction was calculated as:

$$\frac{\textit{number of [middle or high] schools to be selected x individual school's enrollment}}{\textit{total [middle or high] [statewide or regional] enrollment}}$$

W_1, the "school base weight," is the inverse of the sampling fraction. For high and middle schools, W_1 was calculated at the state or regional level as:

$$W_1 = \frac{\textit{total [middle or high] [statewide or regional] enrollment}}{\textit{number of [middle or high] schools to be selected x individual school's enrollment}}$$

School Nonparticipation Factor (F_1)

F_1 is an adjustment factor that accounts for nonparticipation of some selected schools. It was calculated for each sampling stratum as:

$$F_1 = \frac{\textit{number of selected schools}}{\textit{number of participating schools}}$$

Student Selection Probability and Weight (W_2)

For the schools for which all students were selected to participate (i.e., "census"), the student sampling probability is 1. For 22 high schools and 5 middle schools for which a sample of students was selected, the sampling probability was calculated as:

$$\frac{\textit{number of students selected within school}}{\textit{student enrollment within school}}$$

W_2, the student weight, is the inverse of the student sampling probability. For the census schools, W_2 was 1. For the sampling schools, W_2 was calculated at the school level as:

$$W_2 = \frac{\textit{student enrollment within school}}{\textit{number of students selected within school}}$$

Student Nonresponse Factor (F_2)

F_2 is an adjustment factor that accounts for student nonresponse within a participating school and was calculated as:

$$F_2 = \frac{\text{number of students selected}}{\text{number of students who completed the survey}}$$

Post-Stratification Factor within Grade, Sex, and Race/Ethnicity (F_3)

F_3, a post-stratification factor, adjusts for differences between the whole sample and the whole population in the distributions of grade, sex, and race/ethnicity, as reported in the sampling frame. F_3 calculations use student- reported grade, sex, and race/ethnicity. Missing, unreadable, or invalid values on these items were treated as missing at random and imputed based on the joint posterior probability within the school. Imputed values were restricted to appropriate ranges for the school (i.e., grade 7 could not be imputed in a high school).

To construct F_3, a near-final weight was computed for each observation as the product of W_1 x F_1 x W_2 x F_2. Excessively large near-final weights (> [*median* + *{5* × *interquartile range}*]) were trimmed. Near-final weights were summed within each category of state/region-by-grade-by- sex-by-ethnicity and compared to CDE total enrollment for that category. F_3 was computed using raking methods with a result of:

$$F_3 = \frac{\text{total enrollment of students within region-grade-sex-race/ethnicity category}}{\text{total near-final weights within region-grade-sex-race/ethnicity category}}$$

Final Weight

Final weights are the products of base weights and adjustment factors, calculated as:

$$FWT = W_1 \times F_1 \times W_2 \times F_2 \times F_3$$

Response Rates

Response rate (RR) is the product of the school response rate (or participation rate) and the student response rate, calculated as

$$RR = \frac{\text{number of participating schools}}{\text{number of eligible sampled schools}} \times \frac{\text{number of participating students}}{\text{number of sampled students}}$$

A total of 109 sampled middle and high schools and 52,799 sampled students participated in the 2021 state sample HKCS. The overall response rate was 40.3% for the regional/state high school sample and 36.6% for the state middle school sample (Table 2). The response rate for the high school sample is shown by regions in Table 3 (pg. 11).

Table 2: Sample and participation summary, 2021 HKCS

	Sampled Schools	Eligible Sampled Schools	Participating Schools	School RR	Sampled Students	Participating Students	Student RR	Overall RR
High School State Sample	146	145	88	60.7%	68,399	45,363	66.3%	40.3%
Middle School State Sample	40	39	21	53.8%	10,941	7,436	68.0%	36.6%
YRBS High School	40	40	24	60.0%	1,381	1,044	75.6%	45.4%
Ute Tribal concentration*	6	6	3	50.0%	1,252	815	65.1%	32.5%
1801/ Healthy Schools grant*	7	7	6	85.7%	2,105	1,500	71.3%	61.1%

*Some schools are also in the state sample

*Table 3: Response rates by region, 2021 high school HKCS**

Region	Eligible Sampled Schools	Participating Schools	School RR	Sampled Students	Participating Students	Student RR	Overall RR
1	7	1	14.3%	450	317	70.4%	10.1%
2	7	7	100.0%	11,928	7,946	66.6%	66.6%
3	7	6	85.7%	7,029	4,943	70.3%	60.3%
4	7	2	28.6%	1,708	901	52.8%	15.1%
5	7	2	28.6%	85	69	81.2%	23.2%
6	7	4	57.1%	1,094	787	71.9%	41.1%
7	7	6	85.7%	3,760	2,544	67.7%	58.0%
8	7	3	42.9%	481	211	43.9%	18.8%
9	7	5	71.4%	2,925	2,259	77.2%	55.2%
10	6	6	100.0%	1,921	1,422	74.0%	74.0%
11	7	5	71.4%	1,500	1,065	71.0%	50.7%
12	7	6	85.7%	4,870	3,724	76.5%	65.5%
13	7	5	71.4%	1,912	1,430	74.8%	53.4%
14	7	4	57.1%	3,506	2,443	69.7%	39.8%
15	7	3	42.9%	5,007	3,016	60.2%	25.8%
16	7	4	57.1%	1,931	1,431	74.1%	42.3%
17	6	3	50.0%	457	324	70.9%	35.4%
18	7	2	28.6%	3,219	894	27.8%	7.9%
19	7	2	28.6%	1,681	966	57.5%	16.4%
20	7	5	71.4%	5,656	3,655	64.6%	46.2%
21	7	7	100.0%	7,279	5,016	68.9%	68.9%
Total	145	88	60.7%	68,399	45,363	66.3%	40.3%

**Regional data are not publicly available for HSRs with overall response rates <25% (HSRs 1, 4, 5, 8, 18, 19).*

Data Dictionary and Constructed Variables

Separate middle and high school data dictionaries list each original survey question in a single row along with its corresponding, constructed dichotomous (yes/no) variable, named QN_*TOPIC_WORDS*. Dictionary columns show alignment between modules A and B, question wording, response options, variable name, variable format, QN variable name, QN variable label and format, QN variable numerator and denominator definitions, and summary text describing what the variable measures. The same QN is used for survey questions that are identical between high school and middle school questionnaires.

Guide to Data Analysis

Separate HKCS data files are provided for high school and middle school (Table 4). All HKCS data should be analyzed using procedures for complex sampling designs in order to obtain approximately unbiased estimates and error terms (variance, standard error, confidence interval, etc.). Design-based procedures are readily invoked in most analytic packages (e.g., SUDAAN, SPSS, Stata, SAS) and include statements declaring the variables that identify strata, clusters or primary sampling units (PSUs), and weights. Ignoring design specifications in analyses will yield biased estimates and error terms.

Table 4: Dataset names and descriptions

Name of Dataset	**Description**	**Analytic Level**
hkcs2021_hs_public.sasb7dat	High school data	State/Region
hkcs2021_hs_st_rate.sasb7dat	High school sampling rate	State
hkcs2021_hs_region_rate.sasb7dat	High school sampling rate	Region

To analyze high school variables, first consult the data dictionary to determine whether the variable of interest is in the core (both Module A & B), Module A only, or Module B only. The corresponding weight and analytic variables are listed in Table 5.

Table 5: Weighting and analytic variables included in HKCS datasets

Variable	**Description**
HSR	Colorado Health Statistics Region for high school regional estimates
STRATA	Colorado Health Statistics Region for high school state estimates
SCHOOLID	Masked school ID – (typically the CLUSTER or PSU variable)
WT_ST	Final middle school state weight
WT_RGN	Final high school state/region weight, used for variables in both Module A & B
WT_RGN_MODA	Final high school state/region weight, used for variables in Module A
WT_RGN_MODB	Final high school state/region weight, used for variables in Module B

When sampling from a finite population without replacement and the sample size is large in relation to the population (e.g., ≥5%), a finite population correction (FPC) needs to be applied to the formula used to calculate standard error.[1] The sampling fraction for the middle school

[1] Cochran, W. G. (1977). Sampling techniques (3rd ed.). New York: John Wiley & Sons.

sample is relatively small (40 middle schools out of a frame of 555 schools or 7%, near the commonly used threshold of 5%) so the FPC can be ignored. However, the sampling faction for high schools can be high due to the stratification sampling based on 21 health regions, and the FPC should be applied. Secondary data files were provided for high school (*hkcs2021_hs_st_rate.sasb7dat, hkcs2021_hs_region_rate.sasb7dat*) that contains total numbers of eligible schools and sampling rate within each health statistic region (stratum). In some strata with a smaller number of total high schools, it is possible that all or nearly all high schools are selected and participated in the survey. In these strata, the sampling fraction goes to 1, the FPC goes to 0, and the standard error goes to 0. This would inappropriately suggest an exact estimate without considering the variation in the student enrollment during an academic year. The sampling fraction can be truncated at some threshold, for example at 0.7, by inflating the total number of schools in the auxiliary school total file (which do not need to be an integer value). An alternative is to create a new file including the truncated sampling rate (the files *hkcs2021_hs_st_rate* and *hkcs2021_hs_region_rate* include the truncated rate variable _RATE_). This step is recommended for analysis at both statewide and regional levels.

For any survey question, results may be analyzed based on the original variable or its dichotomous counterpart. Dichotomous variables (QN_*TOPIC_WORDS*) are useful for estimating prevalence of behaviors, such as the percentage of students who smoked at all in the past 30 days (QN_SMK_30). Original variables contain more detailed information, such as the percentages of students who smoked on a certain number of days in the past 30 days (SMK_30).

Below is a syntax example using SAS[2] for calculating design-based estimates of (QN_*variable*) on both modules in the high school dataset at the **state level**:

```
proc surveyfreq data={hkcs2021_hs_public rate =hkcs2021_hs_st_rate};strata
strata;
    cluster SCHOOLID;
    weight WT_RGN;
    tables QN_variable;
run;
```

[2] SAS/STAT(R) 9.4 User's Guide, Second Edition

To obtain estimates of demographic groups, domain analysis is used. Domain takes the entire sample into account when estimating variance and is thus more statistically efficient (reduced error margins).

Here is an example SAS procedure to obtain statewide **grade-level** estimates using middle school data:

```
proc surveyfreq data={hkcs2021_ms_public};
    cluster SCHOOLID;
    weight WT_ST;
    tables GRADEMSF*QN_variable /row;
run;
```

District and School Reports

The CSPH Survey Team provides participating schools and districts with reports and does not make these reports available to external parties without school or district permission. Reports include “Frequency Reports” containing all local results for a participating school or district, and “Snapshot Reports” that include comparisons to the region and state (when available).

Opt-in Schools

Schools that were not sampled were eligible to opt in to participate in the HKCS. In a district where some schools were sampled and additional schools opted in, data for all schools were included in the district reports. Thus, if a district has identical boundaries with a region, the district report may include more schools than the regional report. Counting sampled and opt-in schools, a total of 340 public schools and 106,799 students participated in HKCS 2021 (see Table 6).

Table 6: Total participation, including opt-in schools

	Total schools	**Total students**
Opt-in high school participation	91	22,918
Opt-in middle school participation	140	31,082
Total high school participation (opt-in and sampled)	179	68,281
Total middle school participation (opt-in and sampled)	161	38,518
Total middle and high school participation	340	106,799

Limitations

The Healthy Kids Colorado Survey (HKCS) is cross-sectional, meaning it comes from a fixed and particular time point. Causality may not be inferred and additional data can be used to support, or triangulate, findings. Additionally, while the data are weighted to represent students in respective regions and the state, the extent to which bias is minimized is dependent on the response rate – the higher the response rate, the less bias is potentially introduced.

Appendix

Demographic Variables

The following tables indicate variable labels for demographic questions included in the 2021 HKCS, specifically for questions on gender, grade, race and ethnicity, sexual orientation, and gender identity.

Gender

What is your gender identity?

A. Female
B. Male
C. Genderqueer/Nonbinary
D. I do not know my gender identity (questioning)
E. I have a different identity

Response value (GENDER_R21)	Variable for reporting value (GENDER)	Group
A	1	Female
B	2	Male
C	3	Nonbinary
D	4	Not Sure
E	5	Other

Grade (middle school)

In what grade are you?

A. 6th grade
B. 7th grade
C. 8th grade
D. Ungraded or other grade

Response value (GRADE)	Variable for reporting value (GRADE)	Group
A	1	6th grade
B	2	7th grade
C	3	8th grade
D	4	Ungraded or other grade

Grade (high school)

In what grade are you?

A. 9th grade
B. 10th grade
C. 11th grade
D. 12th grade
E. Ungraded or other grade

Response value (GRADE)	Variable for reporting value (GRADE)	Group
A	1	9th grade
B	2	10th grade
C	3	11th grade
D	4	12th grade
E	5	Ungraded or other grade

Race/Ethnicity

What racial or ethnic identity do you most identify with? (Select all that apply.)

A. American Indian or Alaska Native
B. Black or African American
C. East or Southeast Asian
D. Hispanic or Latinx
E. Middle Eastern, North African, or Arab
F. Native Hawaiian or Pacific Islander
G. South Asian
H. White
I. Other

Response value (question)	Response value (RACEID_X)	Variable for reporting value (RACE_ID)	Group
A	RACEID_AIAN = A ONLY	1	American Indian/Alaska Native
B	RACEID_AA = A ONLY	2	Black/African American
C	RACEID_SEA = A ONLY	3	East/Southeast Asian
D	RACEID_HIS = A ONLY	4	Hispanic/Latinx
E	RACEID_ME = A ONLY	5	Middle Eastern/North African/Arab
F	RACEID_NHPI = A ONLY	6	Native Hawaiian/Other Pacific Islander
G	RACEID_SA = A ONLY	7	South Asian
H	RACEID_WHITE = A ONLY	8	White
A	RACEID_OTHER = A ONLY	9	Other
SELECTED 2 OR MORE RESPONSE OPTIONS	SELECTED 2 OR MORE RESPONSE OPTIONS	10	Multi-Racial

Response value (question)	Response value (RACEID_X)	Variable for reporting value (RACE_ID2)	Group
A	RACEID_AIAN = A ONLY	1	American Indian/Alaska Native
B	RACEID_AA = A ONLY	2	Black/African American
C OR G	RACEID_SEA = A OR RACEID_SA = A	3	Asian
D	RACEID_HIS = A ONLY	4	Hispanic/Latinx
E	RACEID_ME = A ONLY	5	Middle Eastern/North African/Arab
F	RACEID_NHPI = A ONLY	6	Native Hawaiian/Other Pacific Islander
H	RACEID_WHITE = A ONLY	8	White
A	RACEID_OTHER = A ONLY	9	Other
SELECTED 2 OR MORE RESPONSE OPTIONS	SELECTED 2 OR MORE RESPONSE OPTIONS	10	Multi-Racial

Sexual Orientation

Which of the following best describes you?

A. Heterosexual (straight)
B. Gay or lesbian
C. Bisexual
D. Asexual
E. I describe by sexual identity some other way
F. I am not sure about my sexual identity (questioning)
G. I do not know what this question is asking

Response value (SEX_ID_R21)	Variable for reporting value (SID)	Group
A	1	Heterosexual (straight)
B	2	Gay or Lesbian
C	3	Bisexual
D	4	Asexual
E	5	Other
F	6	Not sure

Gender Identity

Some people describe themselves as transgender when their sex at birth does not match the way they think or feel about their gender. Are you transgender?

A. No, I am not transgender
B. Yes, I am transgender
C. I am not sure if I am transgender
D. I do not know what this question is asking

Response value	Variable for reporting value	Group
A	1	Cisgender
B	2	Transgender
C	3	Not sure