

# Oregon Healthy Teens 2004 Methodology

## Introduction

The Oregon Healthy Teens Survey (OHT) is the one state-sponsored survey designed to monitor the health and well being of adolescents. This survey reflects the information needs of state agencies and local groups as well as the National Cancer Institute (NCI) funded statewide tobacco prevention evaluation study. An anonymous and voluntary research-based survey, the OHT is designed and administered through a collaborative group of Oregon state agencies including the Department of Human Services, the Department of Education, the Governor's Commission on Juvenile Justice, the Commission on Children and Families, the Oregon Progress Board and the Oregon Research Institute. State and local agencies depend on OHT to assess youth needs, develop comprehensive plans and prevention programs, solicit funding and measure outcomes.

These were the operational goals for the 2004 Oregon Healthy Teens (OHT) survey:

- Ensure voluntary and confidential participation for students, with parental notification;
- Provide a single statewide survey framework encompassing topics of risk behavior and influences on youth behavior for a comprehensive look at youth well-being;
- Draw a stratified random sample of schools across the state in a way that maximized the ability to provide statistically accurate county-level data reporting;
- Obtain a participation rate of at least 60 percent from the randomly selected statewide samples; and
- Allow as many Oregon public high schools and middle schools as possible the opportunity to participate, ensuring our ability to provide localized reports to schools and counties.

If you have any questions please contact the Center for Health Statistics at 503-731-4354 and ask for the Survey Unit Coordinator, the manager of the Health Statistics Unit or the OHT Research Analyst.

## Study Design

### *Initial Process and Instrument Development*

Participation in the OHT was voluntary at every level. District school superintendents were initially contacted late in the fall of 2003 to invite their participation and to request permission to contact their school principals. If district approval was obtained, the school's principal was contacted to obtain approval and the name of a school contact for the survey. Schools and districts received a draft copy of the 2004 questionnaire.

The survey instruments have been revised annually since 1991 and the current focus areas include:

- Tobacco, alcohol and other drug use and access to substances;

- Protective factors and assets such as parental supervision and neighborhood characteristics;
- Physical exercise, nutrition and body weight;
- Sexual risk behaviors;
- Mental health concerns such as suicidal ideation, depression and harassment;
- Intentional and unintentional injury including violence and vehicle safety;
- Health conditions and health care access; and
- Basic demographics.

This year's surveys consisted of one version for 11<sup>th</sup> graders, containing 299 separate response items, and two versions for 8<sup>th</sup> graders, containing either 243 or 266 response items with 64% of the items common to both forms. The survey forms are available on the OHT page at <http://egov.oregon.gov/DHS/ph/chs/youthsurvey/ohtdata.shtml>.

### *Survey Administration*

There were two different groups who performed the actual survey scheduling, collection and initial data processing. The Oregon Research Institute (ORI) has the NCI funding for the 8<sup>th</sup> grade tobacco prevention program evaluation and conducted the survey for those schools covered by their funding, which amounted to about 24% of the 8<sup>th</sup> grade surveys received. For the rest of the survey, the Oregon Department of Human Services (DHS) contracted with a private agency, NPC Research, to perform these tasks. There were minor differences in the survey administration protocol between these two groups but since they used the same survey instruments, and generally the same administration procedures, the data obtained by either group are equivalent.

Surveys conducted by NPC Research used an active notification, passive consent permission model. At least two weeks prior to the scheduled survey date, participating schools sent parents/guardians a letter, notifying them about the survey and asking permission for their child to fill out the questionnaire. Copies of the survey were available at the school and on the Internet for parents and others who wished to have more information about survey content. Parents/guardians were asked to contact the school by mail or phone if they wished to **refuse** consent. Parents could also tell their children not to take the survey and students could also opt out of the survey on their own, even if their parents had not explicitly asked them to be excluded. NPC Research staff provided instructions and technical consultation to the school personnel in each building who coordinated the survey. The surveys were generally administered in the classrooms and students who did not participate were provided with an alternative activity.

### *Student Confidentiality and Privacy*

School staff that administered the survey were required to read and sign confidentiality forms. Students were informed that the survey was anonymous and that they could choose whether to participate and that if they did participate, they could skip any question they did not want to answer. Students completed the questionnaires (which contained no personal identifiers) that were then placed in an envelope. The last student to complete the survey within a classroom was asked to close and seal the envelope and write their name across the seal. Students not

participating in the survey were provided with an alternative activity by their school outside of the classroom.

Only aggregated reports were sent to the schools and districts participating, ensuring student confidentiality. (See reporting for further report dissemination details.) These reports were broken down by grade and gender, except reports were not divided by gender if there were fewer than 10 males and 10 females in the school's valid sample.

### ***Sampling***

The study target population is all Oregon 8<sup>th</sup> and 11<sup>th</sup> grade students. There were three ways schools could enter into the survey; 1) As part of a statewide random sample; 2) As part of the Oregon Research Institute (ORI) Tobacco Prevention evaluation (8<sup>th</sup> grade students only); or 3) As a volunteer school.

As noted in the introduction, the sampling strategy was to not only obtain a representative statewide sample, but also to select this stratified random sample in such a way that also maximized the ability to provide statistically accurate county-level data reporting. The final sample was targeted for approximately 30% of the state's 8<sup>th</sup> and 11<sup>th</sup> grade students, amounting to about 27,000 surveys across 240 schools.

The initial step was to stratify by county, then for each county the following was done:

- Specify the number of districts,  $x$ , to be drawn within each county as proportional to the number of students in the county. A minimum of two districts were chosen, except in counties with only one district. For larger counties a larger number of districts were selected.
- Randomly sample districts proportional to their enrollment; the district is the primary sampling unit.
- The majority of districts have one high school and one or more feeder middle schools. In sampled districts with more than one high school, randomly sample half of the high schools in the district.
- In schools with enrollments of more than 240 in a grade, the number of surveys at that school was limited to approximately 200 students, through random selection of 8-10 classrooms.
- Replacements for districts refusing to participate were to be drawn from the district adjacent in size (alternating between smaller/larger) from among those not selected in the county. If no replacement is available in the county, then a replacement may be drawn from adjacent districts in a DHS Service Delivery Area (SDA) grouping. (See paragraph below.)
- Survey 8<sup>th</sup> and 11<sup>th</sup> graders in each selected high school and associated feeder middle schools.

There were some exceptions to this strategy for the fifteen counties with a single school district, or the majority of students in one district, therefore these districts were selected with certainty. Each of these counties contains less than 1% of the statewide student population and the counties

are: Baker, Crook, Curry, Gilliam, Grant, Harney, Hood River, Jefferson, Lake, Morrow, Sherman, Tillamook, Union, Wallowa and Wheeler.

The final sample totaled 99 high schools and 119 middle schools with the following table listing the school counts by statewide random sample, ORI tobacco evaluation sample, a combination of these two or volunteer schools:

8th Grade				11th Grade	
OHT random sample	Both random and ORI	ORI only sample	Volunteer	OHT random sample	Volunteer
86	18	4	11	90	9

Surveys were returned from 24,425 students, representing an overall response rate of 79.5% of those sampled. Of this total, 2.5% were excluded because of extensive patterns of discrepant and/or dubious (extreme) answers, 3.7% were excluded because their grade level could not be determined, and a further 0.7% were excluded because of missing gender information. (See the response validity section below for more information on this process.) This left 22,750 valid surveys (93.1% of the total received, for a **valid** response rate of 74%), with 9,247 from 11<sup>th</sup> grade and 13,503 from 8<sup>th</sup> grade students in 35 counties. There was no data collected at all from Lincoln county. Union county had no 11<sup>th</sup> grade data; Josephine and Sherman counties had no 8<sup>th</sup> grade data collected.

## Response Validity and Data Editing

### *Initial Data Processing*

The surveys were optically scanned by a private contractor, or by the Oregon Research Institute for the surveys they administered. This process created raw data files containing the responses for each survey received, identified only by the imprinted survey number, which cannot be associated with any particular student. Further processing added school, district, county and other demographic information and NPC Research combined all of these into a single SPSS data file combining the results for all the surveys. If a particular survey contained no grade information an attempt was made to impute the true grade based on the survey form used – since there were different forms used by grade – and by their age. This improved the number of usable surveys but there were still 1,016 (4.2%) surveys where grade information was missing and could not be imputed. An additional 170 surveys (0.7%) had grade information but missing gender information. A table at the end of the next section will summarize these factors, as well as additional data editing based on the following validity testing.

### *Do Oregon Teens Tell The Truth?*

Studies indicate that most young people are truthful in answering anonymous health surveys. Most of the Oregon Healthy Teens survey questions come from studies that have demonstrated

good test-retest reliability in prior research<sup>1</sup>. While a small number of participants did misrepresent their true behavior, the most egregious examples are not included in these data. Specific database processing language was created to verify the validity of responses, including checking for inconsistent and/or dubious answers.

Inconsistencies include things like reporting having smoked cigarettes in the past 30 days, yet reporting never having smoked when asked age of first use. If a survey showed a relatively large number of inconsistencies, the entire survey was marked as invalid. The threshold for this invalidation was 10% or more of the total possible inconsistencies in the survey. If total inconsistencies were less than 10%, the discrepant items were resolved by setting to missing the responses for the inconsistent questions. In general, the first item of the related set was used as the standard, i.e., no change was made to the answer in the standard, or key indicator. Then if the student had inconsistent answers in related questions, and their survey was still within the 10% overall validity threshold, the inconsistent answers in the set of related questions were changed to missing.

Dubious responses are excessive risk behaviors reported by a student. While it could be expected that the highest risk students would report multiple risk behaviors, it is also possible to weed out students who did not take the survey seriously by looking at how many extreme answers in a row were chosen. This dubious response analysis grouped items within topic areas:

- Alcohol, tobacco and other drug use and access;
- Sexual risk behavior;
- Physical activity and nutrition;
- Safety and injury related behaviors; and
- Mental and physical health.

Variables were created to count outlier answers within these topic areas. In a similar fashion to how inconsistent responses were processed, if there were sufficiently large numbers of dubious responses, the entire survey was marked as invalid. The threshold for invalidating the survey for this reason was 20% or more of the total possible dubious responses. If a survey had less than 20% dubious responses overall, but a particular series of questions showed dubious responses, the data for that series of questions was marked as missing.

Of the 24,425 surveys from participating 8<sup>th</sup> and 11<sup>th</sup> graders, 597 (2.5%) were excluded based on validity criteria relating to inconsistent and/or dubious response patterns. (Of these, 108 also had missing grade information.) An additional 1,078 (4.4%) were excluded solely for missing grade or gender information, leaving 22,750 valid surveys, 93.1% of all received. This table summarizes the number of records eliminated by the various editing criteria:

---

<sup>1</sup> Brener ND, Collins JL, Kann L, Warren CW, Williams BI. Reliability of the Youth Risk Behavior Survey questionnaire. American Journal of Epidemiology, 1995; 141:575-80.

	Unknown Grade		8 <sup>th</sup> Grade		11 <sup>th</sup> Grade		Total	
	Count	%	Count	%	Count	%	Count	%
Eliminated by:								
Either inconsistent or dubious criteria	80	7.9%	261	1.9%	194	2.0%	535	2.2%
Both inconsistent and dubious criteria	28	2.8%	23	0.2%	11	0.1%	62	0.3%
Missing gender or unknown grade	908	89.4%	106	0.8%	64	0.7%	1,078	4.4%
<b>Total Eliminated Surveys*</b>	<b>1,016</b>	<b>100.0%</b>	<b>390</b>	<b>2.8%</b>	<b>269</b>	<b>2.8%</b>	<b>1,675</b>	<b>6.9%</b>
<b>Total Usable Surveys*</b>	<b>0</b>	<b>0.0%</b>	<b>13,503</b>	<b>97.2%</b>	<b>9,247</b>	<b>97.2%</b>	<b>22,750</b>	<b>93.1%</b>
<b>Total Received**</b>	<b>1,016</b>	<b>4.2%</b>	<b>13,893</b>	<b>56.9%</b>	<b>9,516</b>	<b>39.0%</b>	<b>24,425</b>	

\* These percentages are of the total number of surveys received within that grade level.

\*\* These percentages are of the total number of surveys received overall.

### ***Weighting***

The surveys from schools that were part of the statewide random sample (11,829 8<sup>th</sup> graders, 8,386 11<sup>th</sup> graders) are weighted, based on various factors. Because the sampling design selected all counties and then schools at random within counties, there are two different weighting values. The first value is to make the sample be representative of all 8<sup>th</sup> or 11<sup>th</sup> grade students within a particular county and the second value is to make the sample be representative of the state. Schools that are not part of the random sample (the 4 tobacco program evaluation only schools and the 20 volunteer schools) had their information reported to their districts but are not weighted and are not included in the DHS county or state reports.

In order to calculate the final weight values, three factors were considered – the base weight for the probability of a particular student being selected within a county, an adjustment for student and school response rates within each county, and a final adjustment for incomplete county coverage. All school enrollment figures are based on Oregon Department of Education enrollment data.

Base Weight. OHT-sampled high schools were sampled within counties proportional to their size and therefore are self-weighting with respect to enrollment. For the purposes of weight calculations, the replacement schools were considered as random with regard to school enrollment size, and were included as simple self-weighting replacements. The base weight was a constant within each county and was equal to the reciprocal of the ratio of the number of students selected per county-grade to the number enrolled in that county-grade.

Response Adjustment. Not all selected schools and/or students within a county participated, therefore the base weights need to be adjusted to reflect the actual participation. These response adjustments were calculated in two parts, using the reciprocal of: 1) The ratio of the number of participating schools to the number selected (school non-response); and 2) The ratio of the number of participating students per school-grade to the number enrolled (student non-response).

County Coverage Adjustment. Because not all counties participated, a final adjustment was necessary to accurately reflect the statewide population. This was done by calculating a post

stratification adjustment as the ratio of total number enrolled per grade in the state to the total grade enrollment of the selected sample.

Final weight. The final weights are the product of base weights, the response adjustments and, for the statewide weight, the county coverage adjustment. The final weights yield county or state population totals per grade.

The published reports for counties and states include weighted percentages for prevalence estimates, but often also include unweighted counts so readers know how many surveys make up each response. Because of the weighting, any percentage recalculation based on the unweighted counts may differ slightly from the published percentages. The published percentages accurately represent state prevalence with a maximum margin of error (95% confidence interval) of approximately  $\pm 0.7\%$  for the 8<sup>th</sup> grade and  $\pm 0.9\%$  for the 11<sup>th</sup> grade. The maximum margin of error occurs with prevalence estimates at or near the 50% level. The exact margin of error for each question is smaller and varies, depending on the percentage estimated for that particular question. The margins of error for county estimates are slightly larger; depending on the numbers of surveys collected in the county, but are generally smaller than  $\pm 5\%$ .

## **Reporting**

### ***Schools***

Summary and detailed reports were sent to each participating school and district and these reports are available from the school district. The summary reports included key indicators around tobacco, alcohol and other drug attitudes and use, school safety, mental health, sexual activity and health-promoting behaviors and these were sent directly to the schools by the DHS contractor or by ORI for schools in the tobacco prevention program study. After the summary reports were sent, DHS prepared detailed reports for each school and district, containing counts and percentages for every survey item. These reports were burned onto CDs and also sent to each school and district.

Because individual schools may have small enrollments, caution should be used when interpreting percentages based on a small number of respondents. Further, the schools were not selected to provide a reliable estimate for the district so caution should be used combining school data within a district. No personally identifiable information is available because the survey was anonymous. To further protect confidentiality, reports were not broken down by gender if there were fewer than 10 males and 10 females in the school's valid sample.

### ***DHS State and County Reports***

OHT county-level and statewide results are available on the OHT website at <http://egov.oregon.gov/DHS/ph/chs/youthsurvey/ohtdata.shtml>. Tabulations by racial and ethnic groupings are also available at this website. The weighted percentages in the race and ethnicity reports were based on the overall statewide sample weights and were not readjusted to reflect enrollment variances by racial or ethnic composition within the state.