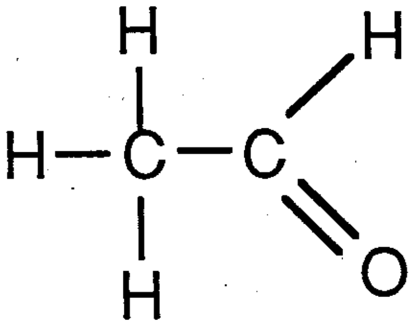
California Environmental Protection Agency

Air Resources Board

Acetaldehyde

as a Toxic Air Contaminant



Part A

Exposure Assessment

Stationary Source Division

November 1993

ACETALDEHYDE AS A TOXIC AIR CONTAMINANT

PART A

TECHNICAL SUPPORT DOCUMENT
PUBLIC EXPOSURE TO SOURCES AND EMISSIONS OF ACETALDEHYDE IN CALIFORNIA

Principal Investigators
Linda Martz
Barbara Cook

Contributors
Vincent Agusiegbe, Steve Brisby, Richard Hackney,
Chris Nguyen, Elizabeth Ota

Dr. Roger Atkinson
University of California, Riverside

Reviewed by
Joan E. Denton, Ph.D., Manager
Substance Evaluation Section

Genevieve A. Shiroma, Chief
Toxic Air Contaminant Identification Branch

Donald J. Ames, Assistant Chief
Stationary Source Division

Peter D. Venturini, Chief
Stationary Source Division

November 1993

ACKNOWLEDGMENTS

The Stationary Source Division acknowledges the valuable assistance of the following people who contributed to the preparation and review of this document: Michele Burkett, Faye Palmer, Doug Lawson, Eric Fujita, Peggy Jenkins, and Michael Poore. We also thank and acknowledge the staff who dedicated their time in the research and written portions of the report, and to the staff who conducted the field work including the monitoring, analysis, and methods development. In preparing this report, staff reviewed pertinent literature through March 1992.

November 1993

REPORT SUBMITTED TO THE SCIENTIFIC REVIEW PANEL

Part A- Public Exposure To Sources and Emissions of Acetaldehyde in California

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

I.

INTRODUCTION

This report summarizes the emissions, concentrations, exposure, and atmospheric persistence and fate of acetaldehyde in California. It also summarizes the health effects of acetaldehyde which includes an estimate of cancer potency.

This report was developed in response to provisions of AB 1807 (Health and Safety Code sections 39650-39662), which became effective in January 1984. This legislation requires a two-phase process which separates risk assessment (identification) of toxic air contaminants (TACs) from risk management (control). During the identification phase, a report is developed which considers whether there are adverse health effects of a substance which may be, or, is emitted in California. After conducting a public hearing, the Board decides whether or not the substance should be identified as a TAC. If identified, the substance is listed by regulation as a TAC in California Code of Regulations and enters the control phase.

With the adoption of the AB 2728 legislation (signed by the Governor in September 1992 and effective in January 1993), the procedure for identifying substances already classified as federal hazardous air pollutants (HAPs) as TACs was changed. Pursuant to the new legislation, the state board identified as TACs on April 8, 1993, substances listed as federal HAPs through a simplified process. Acetaldehyde is a HAP and, therefore, was identified as a TAC under the new process.

After the Board hearing, where all HAPs were identified as TACs, the cancer unit risk numbers for acetaldehyde developed by the Office of Environmental Health Hazard Assessment (OEHHA), were approved by the Scientific Review Panel (SRP) on May 12, 1993. The role of the SRP is to review the data, assessments and conclusions presented in the report as the basis for approving the potency number. The potency number will be used in the control phase of the AB 1807 process. It may also be used by the local districts for permitting decisions, and to assess the risk to public health in the AB 2588 "Hot Spots" program.

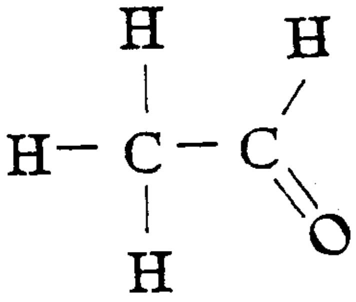
A. ACETALDEHYDE AS A CHEMICAL COMPOUND

Acetaldehyde is an aliphatic aldehyde: a saturated hydrocarbon with a terminal carbonyl group (C=O). Acetaldehyde is known by a variety of synonyms which include: acetic aldehyde, ethanal, ethyl aldehyde and methyl formaldehyde [Chemical Abstracts Service (CAS) Registry Number 75-07-0]. Acetaldehyde is not known to contribute to either global warming or depletion of the ozone layer.

Gasoline containing the oxygenate additives ethanol or ethyl tert-butyl ether (ETBE) upon combustion results in increased acetaldehyde emissions. These oxygenates may be used as additives to meet the requirements of the CARB Phase 2 Reformulated Gasoline Regulations and Wintertime Oxygenates Program which may increase acetaldehyde emissions.

The major acetaldehyde decomposition products are formaldehyde and peroxyacetyl nitrate (PAN) both of which are of concern as toxic species. In neither case is acetaldehyde the dominant source of these species (EPA, 1993).

Acetaldehyde has the molecular formula C_2H_4O and the chemical structure shown here.



B. References for Chapter I

EPA (1993) "Motor Vehicle-Related Air Toxics Study", Technical Support Branch, Emission Planning and Strategies Division, Ann Arbor, MI.

II.

PROPERTIES OF ACETALDEHYDE

A. PHYSICAL PROPERTIES

Synonyms for acetaldehyde are acetic aldehyde, ethanal, ethyl aldehyde and methyl formaldehyde. Acetaldehyde is a saturated aldehyde with the chemical formula CH_3CHO. It is a colorless liquid, volatile at ambient temperature and pressure. Although it has a pungent irritating odor, at dilute concentrations it has a fruity and pleasant odor. The threshold odor concentration in air is 7.8 to 33.3 ppbv (0.014 mg/m^3 to 0.06 mg/m^3). Both the liquid and the vapors are highly flammable (Cooke, 1971) and acetaldehyde is a dangerous fire hazard when exposed to heat or flame (Sax, 1975). As a liquid it is lighter than water and the vapors are heavier than air. Acetaldehyde is soluble in water, acetone, gasoline, alcohols, toluene, xylene, benzene, ether, paraldehyde and organic solvents. It decomposes at temperatures above 400°C (Kirk-Othmer, 1980). The decomposition products are principally methane and carbon monoxide. Acetaldehyde undergoes photochemical decomposition when irradiated by ultraviolet light and some of the products are PAN and formaldehyde. The physical and chemical properties of acetaldehyde are listed in Table II-1.

Acetaldehyde is an intermediate product in the respiration of higher plants and can be found in ripening fruit such as apples. Also, acetaldehyde is an intermediate product of fermentation of alcohol and the metabolism of sugars in the body.

Table II-1

PHYSICAL PROPERTIES OF ACETALDEHYDE

Properties	Value	Reference
Conversion factor: 1 ppm = 1.8 mg/m^3 at 760 mm Hg and 25° C		
Chemical formula	CH_3CHO	
Cas Registry number	75-07-0	a
Molecular weight	44.05	b
Boiling Point, 1 Atm	21 $^{\circ}$C	b
Vapor Density (air = 1)	1.52	a
Melting Point	-123.5 $^{\circ}$C	b
Flash pt. closed cup	-38	b
Density (specific gravity at $18^{\circ}C/4^{\circ}C$)	0.79	a
Dissociation constant (at 0°C, Ka)	0.7×10^{-14}	a
Refraction index (n_D^{20})	1.33113	a
Partition coefficient (Log P octanol/water)	0.43	a
Vapor pressure at 20°C	755 mm Hg	a
Autoignition temperature	193°C	c
Coefficient of expansion per $^{\circ}$C	0.00169	a
Surface tension at 20°C,mN/m	21.2	a
Explosive limits of mixtures with air, vol % acetaldehyde	4.5-60.5	a
Ionization potential,eV	10.50	a
Critical temperature, $^{\circ}$C	181.5	a
Critical pressure, MPa (atm)	6.40 (63.2)	a
Latent heat of vaporization, kJ/mol	25.71	a

a Kirk and Othmer, 1980

b Merck Index, 1983

c WHO, 1985

Acetaldehyde is highly reactive, exhibiting the general reactions of aldehydes; under proper conditions, the oxygen or any hydrogen can be replaced (Kirk and Othmer, 1980). Further, reaction with oxygen may lead to explosion (Sax and Lewis, 1989). It is a strong reducing agent and undergoes numerous condensation, addition and polymerization reactions. In the presence of catalysts such as trace metals or acids, acetaldehyde rapidly polymerizes to form paraldehyde (Fairhall, 1957). Acetaldehyde can react violently with acid anhydrides, alcohols, ketones, phenols, ammonia, hydrogen cyanide, hydrogen sulfide, halogens, phosphorus, isocyanates, strong alkalis and amines (Sax and Lewis, 1989).

Because of its high chemical reactivity acetaldehyde is used as an intermediate in the production of acetic acid, acetic anhydride, ethyl acetate, peracetic acid, pentaerythritol, chloral, glyoxal, alkylamines and pyridines (Kirk and Othmer, 1980).

B. References for Chapter II

Cooke, W.G. (1971) "Acetaldehyde", Encyclopedia of Occupational Health and Safety, Volume 1, New York, McGraw-Hill Book Company, pp. 23-24.

Fairhall, L.I. (1957) "Carbon Compounds: Acetaldehyde", Industrial Toxicology, 2nd ed. Baltimore, MD, Williams and Wilkins, pp. 138-140.

Kirk, R. and D. Othmer (1980) Encyclopedia of Chemical Technology, 3rd Edition, Volume 1, John Wiley and Sons Inc., New York, NY, pp. 97-99.

Merck Index (1983) An Encyclopedia of Chemicals, Drugs and Biologicals, Tenth Edition, Merck & Co., Inc., p. 6.

Sax , N.I. (1975) Dangerous Properties of Industrial Materials, 4th ed., New York, NY, Van Nostrand Reinhold company, p. 348.

Sax, N.I. and R.J. Lewis, Sr. (1989) Dangerous Properties of Industrial Materials, 7th ed., New York, NY, Van Nostrand Reinhold company, p. 6.

World Health Organization (WHO) (1985) "Allyl Compounds, Aldehydes, Epoxides and Peroxides", IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, Volume 36, pp. 101-102.

III.

PRODUCTION, USES AND EMISSIONS OF ACETALDEHYDE

Acetaldehyde is both directly emitted into the atmosphere as well as formed in the atmosphere as a result of photochemical oxidation of organic pollutants in urban atmospheres. Photochemical oxidation is the largest source (could be as high as 67 percent) of annual concentrations in the outdoor ambient air in California. Mobile and stationary sources emit reactive organic gases such as ethyl, ethyl peroxide, and ethoxy radicals which are precursors of photochemically generated acetaldehyde. Reductions of these hydrocarbon precursors can be expected to reduce the contribution of acetaldehyde from photochemical oxidation.

The direct release of acetaldehyde from combustion processes include vehicular exhaust from mobile sources and fuel combustion from stationary internal combustion (IC) engines, boilers and process heaters. Combustion of gasoline containing the oxygenate additives ethanol or ETBE may result in increased acetaldehyde emissions. Other sources of acetaldehyde include refineries, coffee bean roasters, residential wood combustion, wildfires, agricultural burning and management burning.

A. PRODUCTION

Acetaldehyde is produced two different ways: formation by photochemical oxidation in the atmosphere and commercial manufacture. In California, photochemical oxidation is the major source of acetaldehyde formation.

1. Photochemical Contribution

Although several kinetic mechanisms for the formation of acetaldehyde in the atmosphere have been proposed (Finlayson-Pitts, B.J. and J.N. Pitts, Jr., 1986; and Atkinson et al., 1990), as of this writing, there are no published estimates of the annual concentrations of acetaldehyde from photo-oxidation formation in the atmosphere. As an approximation, the ARB staff performed an Urban Airshed Model (UAM) simulation using ambient data for August 28, 1987, in the South Coast Air Basin (SoCAB) to estimate the portion of the ambient concentration of acetaldehyde due to secondary formation. The UAM simulates the ambient acetaldehyde concentrations from both direct and secondary formation (Ligocki and Whitten, 1991). Because August 28, 1987, was a summer high-ozone day, the staff believes that this simulation overpredicted the annual average secondary acetaldehyde. Based on this simulation, the staff estimates that the worst-case SoCAB average ambient surface concentration of acetaldehyde in the atmosphere resulting from secondary formation ranges from approximately 41 percent during the night to 67 percent during the day of the total acetaldehyde concentration in the atmosphere (Wagner, 1991). If the same location and time of emissions in the UAM model are assumed, then from the emission inventory in Table III-1 a ratio can be used to approximate the statewide secondary acetaldehyde component (Appendix A 2). We estimate that the secondary acetaldehyde is equivalent to a range of 14,000 to 31,000 tons of primary acetaldehyde emissions annually.

2. Commercial Production

As of 1990, there are two acetaldehyde producers in the United States; neither is located in California (SRI International, 1989). Acetaldehyde is also produced as a by-product in the production of polyesters, dimethyl terephthalate, acetone, mixed esters, ethanol, ethyl ether, vinyl acetate, ethylene glycol and acrolein. None of these facilities is located in California [Midwest Research Institute (MRI), 1987]. However, there are seven producers of polyester resins in California (SRI International, 1989). Information is not available to assess whether these facilities emit acetaldehyde.

Table III-1

SUMMARY OF CALIFORNIA ACETALDEHYDE EMISSIONS IN 1987

Source	Emissions* (Tons Per Year)	Notes (see following page)
PRODUCTION		
As a Photo-oxidation Product	14,000-30,000	a
EMISSIONS		
Mobile Sources		
On-Road Vehicles	1,600-5,700**	b,c,d,e
Other Mobile Sources	2,100	c,d,e
Stationary Area Sources		f
Residential Wood Combustion	110-660	
Natural Gas Combustion	7-14**	
Diesel Combustion	43	
Oil Combustion	1-2	
Agricultural Burning	1,000-2,100	g
Management Burning	1,200-2,500	g
Open Burning	87-180	g
Wildfires	4,500-9,200	g
Stationary Point Sources		h
Fuel Combustion		
Coal	1-2	
Diesel	1	
Gasoline	10	
Natural Gas	15-31	
Process Gas	***	
Oil	10-20	
Wood	560	
Refineries - Processing	190-770	i,j
Coffee Bean Roasting	2	k
Food Preparation	7	l

* Emissions have been rounded to, at most, two significant figures.

** A small portion (less than 1%) of this estimate is for activities occurring in the outer continental shelf area.

*** Acetaldehyde emissions from process gas combustion may equal approximately 50 percent of the estimated emissions from natural gas combustion.

Notes:

a. Atmospheric formation of acetaldehyde varies widely during the period of day, meteorological conditions and geographical location. Estimates can vary by 50 percent. Using data from the the Urban Airshed Model it is estimated that 41 to 67 percent of acetaldehyde is photochemically formed in the atmosphere (see Appendix A). In the UAM simulations, the model species ALD2 is assumed to approximate acetaldehyde.

b. ARB (1990a)

c. ARB (1990b)

d. ARB (1990c)

e. ARB (1989)

f. See Appendix B, Stationary Area Sources, Table 6

g. These combustion sources may contribute a relatively low exposure concentration due to lack of proximity to significant populations and high plume rise.

h. See Appendix B, Stationary Point Sources

i. Oil and Gas Journal (1989)

j. See Appendix B, Stationary Point Sources, Refineries

k. ARB (1990i)

l. SARA 313 data

B. USES

Acetaldehyde is used in the synthesis of many chemicals such as acetic acid, paraldehyde, pyridines, terephthalic acid, pentaerythritol, perfumes, plastics, 1,3-butalene glycol and synthetic rubber. It is also used as an industrial solvent and as a food flavoring agent preservative (Serth et al., 1978). None of the above mentioned chemical manufacturers is located in California. No information is available on the use of acetaldehyde as an industrial solvent or as a food flavoring agent preservative.

C. EMISSIONS

Table III-1 summarizes acetaldehyde emissions from all accountable sources in California. Figures III-1 and III-2 show the percentages of acetaldehyde emissions from the listed sources. Mobile sources which include on-road motor vehicles and other mobile sources such as trains, ships, farm and utility equipment emitted approximately 3,700 tons to 7,800 tons of acetaldehyde in 1987. Stationary point sources such as fuel combustion sources, refineries and coffee bean roasters emitted approximately 820 to 1,000 tons of acetaldehyde in 1987.

Stationary area sources are those stationary sources not otherwise included in the point source category. These sources include diesel combustion in internal combustion engines at oil and gas fields, wildfires, agricultural burning and management burning, etc. The staff estimates that these stationary area sources emitted approximately 7,000 to 15,000 tons of acetaldehyde in 1987. The following sections discuss each category in detail.

1. Mobile Sources

a. On-Road Motor Vehicles

The staff estimates that on-road motor vehicles including light-duty passenger cars, medium-duty trucks, heavy-duty trucks, buses and motorcycles emitted approximately 1,600 to 5,700 tons of acetaldehyde in 1987 (see Appendix B). Other transportation sources emitted approximately 2,100 tons. The lower limit was estimated based on the emission factors from several studies and the total organic gas (TOG) emissions from vehicular exhaust (ARB, 1990a) while the upper limit was estimated based on the volatile organic compound (VOC) speciation profiles and the TOG emissions in the ARB's Emission Data System (EDS) (ARB, 1990b and ARB, 1990c).

Figure III-1
Sources of Acetaldehyde for 1987
in California

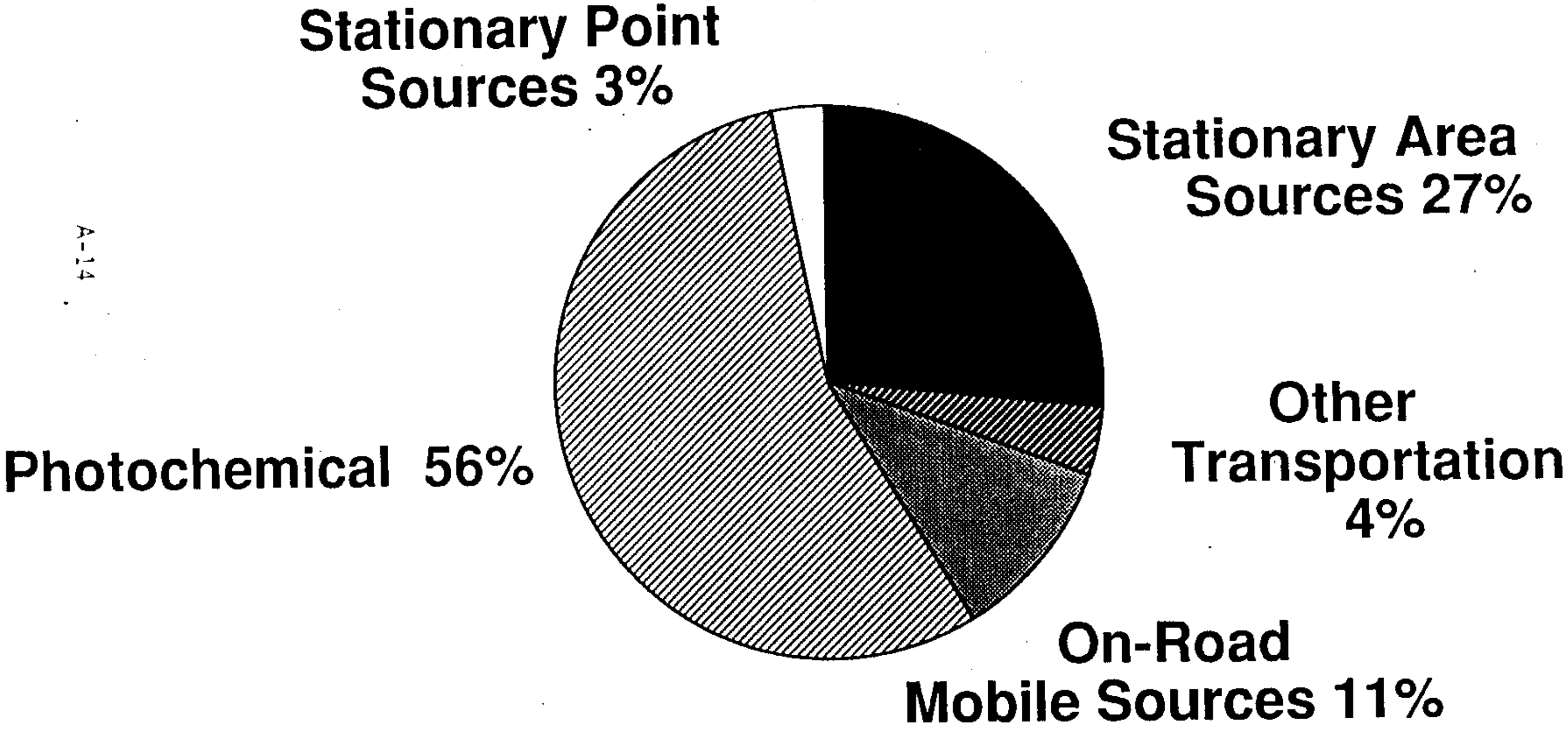
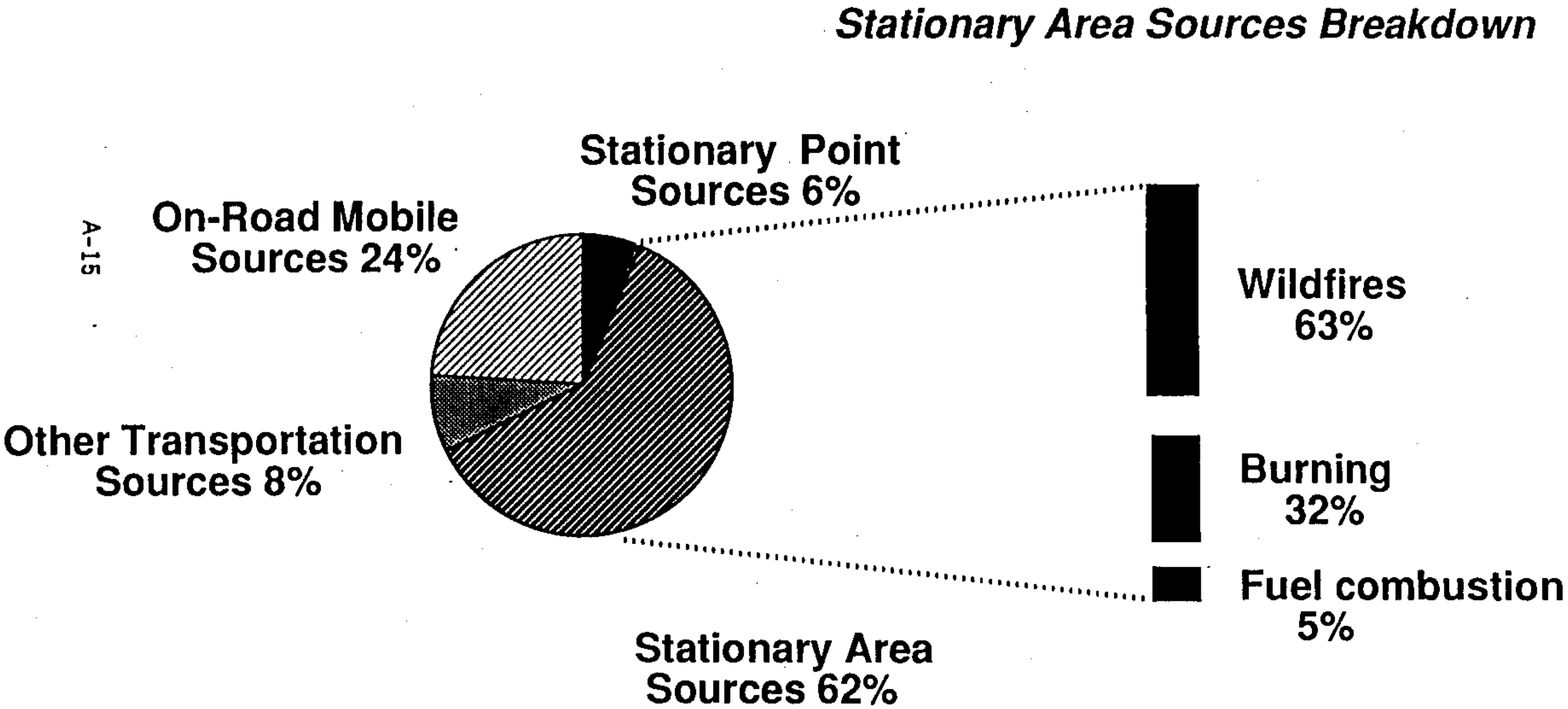


Figure III-2
Direct Sources of Acetaldehyde
for 1987 in California



During the past few years, a number of independent investigators have conducted studies that show that the total hydrocarbon emission inventory (representing all types of sources) may be underestimated by substantial amounts. Investigations conducted during the last year by the ARB staff have shown that these underestimates are in the neighborhood of 50 to 100 percent. ARB staff believes that a significant portion of this error is in the on-road motor vehicle portion of the inventory; however, studies to date have not been able to establish error bands for specific categories of the inventory. Efforts towards improving both the mobile and stationary source portions of the inventory continue and a major effort is underway to obtain improved emission rates and vehicular activity data for the on-road motor vehicle emission estimates.

Acetaldehyde and other aldehydes were studied as non-regulated chemicals from automotive exhaust (Carey, 1981; Urban, 1981; Urban, 1980a; Urban, 1980b; Urban, 1980c; Sigsby, Jr. et al., 1987; Springer, 1979; Ullman, and Hare, 1985). The absolute acetaldehyde emission factors in milligrams per kilometer and the total hydrocarbon (THC) emission factors in these studies were used to normalized acetaldehyde as fractions of THC. The staff used these fractions along with the ratios of THC to TOG emissions from motor vehicles in California (ARB, 1988) to estimate the lower limit of acetaldehyde emissions.

The ARB has also compiled VOC speciation profiles for on-road motor vehicles. These profiles express specific substances as a weight fraction of TOG. The staff used acetaldehyde fractions in these profiles and the TOG emissions from on-road vehicular exhaust to estimate the upper limit of acetaldehyde emissions.

Lower and upper limits of estimated acetaldehyde emissions for each of the vehicular classes are reported in the following paragraphs.

Non-catalyst, light-duty passenger cars emitted approximately 500 to 2,300 tons of acetaldehyde while those equipped with catalytic converters emitted approximately 390 to 820 tons in 1987. For those light-duty passenger cars that burn diesel fuel, the acetaldehyde emissions were approximately 8 to 54 tons.

Light and medium-duty gasoline trucks equipped with catalytic converters emitted less acetaldehyde (approximately 170 to 350 tons) when compared to those not equipped with catalytic converters (about 180 to 810 tons) in 1987. In the same inventory year, the staff estimates that light and medium-duty diesel trucks emitted approximately 4 to 14 tons of acetaldehyde.

Acetaldehyde emissions from all heavy-duty gasoline trucks (non-catalytic) were significantly higher (approximately 94 to 390 tons) than acetaldehyde emissions (about 1 ton) from heavy-duty diesel trucks equipped with catalytic converters. The staff estimates that heavy-duty diesel trucks emitted approximately 280 to 870 tons of acetaldehyde in 1987.

Urban buses burning diesel emitted approximately 12 to 37 tons of acetaldehyde in 1987. In the same inventory year, motorcycles were responsible for approximately 22 to 100 tons of acetaldehyde emissions.

b. Other Mobile Sources

Acetaldehyde emissions associated with all off-road motor vehicles and other mobile sources total about 2,100 tons for 1987 (see Appendix B). Of this total, trains accounted for 320 tons, while ships and aircraft accounted for 110 tons. Off-road motor vehicles such as recreational vehicles and commercial boats were responsible for approximately 780 tons. Mobile and utility equipment such as tractors, refrigeration units, fork lifts and lawnmowers accounted for approximately 870 tons of acetaldehyde (see Appendix B).

2. Trends

Most of the contribution by motor vehicles to the ambient burden of primary acetaldehyde is from the exhaust of gasoline vehicles. They also contribute to the ambient burdens of oxides of nitrogen (NOx) and reactive organic gases (ROG), which react via photosynthesis in the atmosphere to create secondarily formed acetaldehyde.

a. Motor vehicle technology

The emissions standards applied by the ARB to new vehicles over the past years have led to significant reductions in vehicular ROG and NOx directly emitted by motor vehicles. Further, these reductions in ROG and NOx have resulted in a decline of secondary acetaldehyde. Declines in motor vehicle emissions will continue because of the recently adopted "low-emission vehicle" (LEV) standards for non-methane organic gases and NOx. Additionally, the primary, directly emitted acetaldehyde, also a ROG, is expected to decline. This conclusion is supported by the Auto/Oil data (Auto/Oil, 1991) and ARB studies (ARB 1991c, ARB 1990d, ARB 1990e, ARB 1990f).

b. Gasoline specifications

The ARB adopted two new gasoline regulations, Phase 2 Reformulated Gasoline standards which become effective in 1996 and the Wintertime Oxygenates Program which became effective in 1992. The Phase 2 gasoline standards were designed to achieve reductions in criteria and toxic pollutants and will reduce ROG and NOx emissions from all gasoline vehicles (vehicles currently on the road plus LEVs made before 1996). Reduction of reactive hydrocarbons and NOx will decrease the formation of both primary and secondary acetaldehyde.

The Wintertime Oxygenate Program sets standards for the oxygen content in gasoline sold in California during the winter months in an effort to reduce carbon monoxide emissions. Gasoline oxygenate additives used to provide the minimum oxygen content required by the standards could increase acetaldehyde emissions (Auto/Oil, 1990-1991). The addition of methanol and methyl tert-butyl ether (MTBE) to fuel does not increase acetaldehyde emissions. Alternate fuels containing the oxygenate additives ethanol or ethyl tert-butyl ether (ETBE), upon combustion, result in acetaldehyde emissions. However, it is not known to what extent ethanol or ETBE will be used as a winter oxygenate in California fuels and, therefore, the resulting affect on actaldehyde emissions.

Acetaldehyde increases could elevate peroxyacetyl nitrate (PAN) and formaldehyde concentrations. PAN is one of the main products of the photooxidation of acetaldehyde in the presence of NOx and is a strong eye irritant and plant toxicant. However, acetaldehyde is not the dominant source of PAN as many other hydrocarbon species are important PAN precursors in urban atmospheres (EPA, 1993) (Atkinson, 1993).

The overall effect of the vehicular emission standards and the gasoline standards is a complex issue. The acetaldehyde concentration trend will depend on the turnover rate of older vehicles from the fleet mix to LEVs, the primary acetaldehyde contributions from oxygenate additives, and the decrease of ROG from the Phase 2 reformulated gasoline.

3. Stationary Area Sources

The stationary sources such as internal combustion engines in the oil and gas fields are sources that are not included in the point source category. These sources are scattered throughout the State. Individually, they may not contribute significantly to air pollution; however, in the aggregate, their toxic air emissions are of potential concern. The staff estimates that these sources emitted approximately 7,000 to 15,000 tons of acetaldehyde in 1987 as a result of fuel and waste combustion (see Appendix B).

a. Residential Wood Combustion

Based on data from source tests conducted on several woodstoves and fireplaces, acetaldehyde emission factors range from 0.06 to 0.19 lb/ton of wood burned (MRI, 1987). Using a weighted average of these data and the amount of wood burned for residential purposes (ARB, 1990g), the staff estimates that residential woodstoves and fireplaces emitted approximately 110 tons of acetaldehyde in 1987 (see Appendix B). Alternatively, using the TOG emissions from residential wood combustion and an applicable speciation profile (ARB, 1989), the staff estimates that residential woodstoves and fireplaces emitted approximately 660 tons of acetaldehyde in 1987 (ARB, 1990b).

b. Diesel and Oil Combustion

Stationary area sources of diesel, distillate and residual oil combustion include boilers and IC engines in oil and gas fields and other sources such as orchard heaters in agricultural production. The staff estimates that these combustion sources emitted approximately 44 to 45 tons of acetaldehyde in 1987, primarily as a result of incomplete combustion (see Appendix B).

c. Natural Gas Combustion

In 1987, California used approximately 844 billion cubic feet of natural gas for heating, cooking, etc. (ARB, 1990a). Emissions from these combustion sources may be considered to be a source of indoor air pollution. However, the staff assumed that emissions from these sources escape to the outside. Based on the amount of natural gas burned and the acetaldehyde emission factors developed by MRI from a formaldehyde emission factor, the staff estimates that natural gas combustion sources emitted approximately 7 to 14 tons in 1987 (see Appendix B).

d. Agricultural, Management, Open Fires and Wildfires

The nature of wildfires and management and agricultural fires make it almost impossible to control or predict them even through these fires may emit large amounts of acetaldehyde. The ARB compiles activity and criteria pollutant emission data for these fires. The staff is not aware of any specific acetaldehyde emission factor for wildfires and management and agricultural fires. Based on the aldehyde emission factor from backyard burning, MRI estimated that wildfires, management and agricultural fires emitted approximately 0.73 to 1.5 lbs of acetaldehyde per ton of material burned (MRI, 1987). These emission factors were developed using national data and may not necessarily be representative of California conditions. However, for lack of any better data, the staff used the emission factors developed by MRI and the activity data compiled by the ARB to estimate that agricultural, management, open fires and wildfires emitted approximately 6,800 to 14,000 tons of acetaldehyde in 1987 (see Appendix B).

4. Stationary Point Sources

a. Fuel Combustion

Acetaldehyde is formed from combustion processes as a result of incomplete combustion. The degree of acetaldehyde emissions depends on factors such as the type of fuel burned, the fuel-to-air ratio, the operating conditions and the design of the combustion devices. However, the staff is not aware of any studies relating the combustion parameters to the concentration of acetaldehyde in the exhaust. Based on the amounts of fuel burned in California or the TOG emissions and the emission factors for specific types of fuel, the staff estimates that stationary point sources of fuel combustion (including the combustion of coal, distillate and residual oil, diesel, gasoline, natural gas and wood) emitted approximately 600 to 630 tons of acetaldehyde in 1987.

There is a great deal of uncertainty in the acetaldehyde emission estimates from the combustion of coal, natural gas and residual and distillate oil. In these cases, MRI used formaldehyde emission factors to estimate total aldehyde emissions by assuming that formaldehyde represents 70 percent of total aldehydes. Once total aldehyde emissions were estimated, MRI further assumed that total aldehydes consist of 7 to 14 percent acetaldehyde to estimate acetaldehyde emissions (MRI, 1987). The percentages of acetaldehyde in total aldehydes were derived from source tests of residential wood combustion (MRI, 1987). The following sections discuss acetaldehyde emissions for each fuel type for which quantitative estimates of emissions have been made. Other fuel types for which emissions could not be quantified are discussed in Section 4.

Coal

In 1987, California used approximately 1.6 million tons of coal resulting in 1 to 2 tons of acetaldehyde emissions (see Appendix B). There is no specific acetaldehyde emission factor for coal combustion. MRI used an emission factor for formaldehyde and assumed that formaldehyde represents 70 percent of total aldehydes in the combustion exhaust and that the percentage of acetaldehyde to total aldehydes in coal combustion is the same as that in wood combustion (MRI, 1987). Based on these assumptions, acetaldehyde emission factors for coal combustion range from 0.0014 to 0.0029 lb/ton.

Diesel

The staff estimates that stationary point sources in California emitted approximately 1 ton of acetaldehyde in 1987 as a result of diesel combustion. This estimate was based on the TOG emissions from all California stationary point sources burning diesel and the fraction of acetaldehyde in TOG (ARB, 1990g and ARB, 1989).

Gasoline

Besides motor vehicles, which are a major user of gasoline, stationary point sources also use gasoline. In 1987, these sources used approximately 4.53 million gallons of gasoline and emitted approximately 10 tons of acetaldehyde in California. The staff estimates these emissions using the fraction of acetaldehyde in TOG (ARB, 1989) and the TOG emissions from stationary point sources (ARB, 1990h).

Natural Gas

The staff estimates that natural gas combustion sources in California emitted approximately 15 to 31 tons of acetaldehyde in 1987 (see Appendix B). This estimate was based on the amount of natural gas burned at stationary point sources in California and the estimated emission factor for natural gas combustion developed by MRI (MRI, 1987). This emission estimate is very rough because the emissions were derived from the emissions of formaldehyde to total aldehydes and then backcalculated to acetaldehyde.

Oil

The staff is not aware of any specific acetaldehyde emission factors from distillate or residual oil combustion. For lack of better data, the staff has applied the methodology developed by MRI to estimate acetaldehyde emissions from oil combustion.

In 1987, California burned approximately 3.25 billion gallons (approximately 12.6 million tons) of distillate and residual oil in stationary point sources. Using the methodology developed by MRI, the staff estimates stationary point sources of oil combustion in California emitted approximately 10 to 20 tons of acetaldehyde (see Appendix B).

Wood

The staff has applied an acetaldehyde emission factor from residential wood combustion to estimate emissions from industrial wood combustion. The staff developed this emission factor as a weighted average of several source tests for woodstoves and fireplaces burning different types of woods. Using the data on the amount of wood burned for industrial purpose (ARB, 1990i) and the assumed emission factor, the staff estimates industrial wood combustion sources emitted approximately 560 tons of acetaldehyde in 1987 (Appendix B).

b. Refineries - Processing

Acetaldehyde is a product of incomplete combustion in petroleum refineries. Catalytic cracking, coking operations and fuel combustion are major sources of acetaldehyde from refineries. Other operations such as catalytic reforming, catalytic hydrorefining and catalytic hydrotreating may also be sources of acetaldehyde. However, there is no information regarding aldehyde emissions from these sources.

Catalytic cracking refers to operations where catalysts are used to break down heavy oils (high molecular weight compounds) to lighter products. The spent catalysts are then regenerated by combusting the deposited coke. This combustion process emits acetaldehyde. Coking operations involve the thermal coking of heavy residual oil to other products and petroleum coke (EPA, 1984). Incomplete combustion of coke generates acetaldehyde. Acetaldehyde emissions from fuel combustion in refineries have been estimated in the fuel combustion section. This section only discusses acetaldehyde emissions from petroleum processing.

There is no specific acetaldehyde emission factor for refinery processes. The Science Applications International Corporation (SAIC) assumed in general that 20 percent of total aldehydes is acetaldehyde was based on the percentages of acetaldehyde to total aldehydes in vehicular exhaust from several studies (SAIC, 1987). In 1980 EPA published total aldehyde emission factors for catalytic cracking and coking operations (EPA, 1980). By assuming that acetaldehyde represents 20 percent of total aldehydes from petroleum refinery processing, acetaldehyde emission factors are calculated to be 1.4 lbs per 1,000 barrels (bbls) of fresh feed from fluidized catalytic cracking (FCC); 0.63 lb per 1,000 bbls of fresh feed from moving-bed cracking or thermal catalyst cracking (TCC); and 0.34 lb per 1,000 bbls of fresh feed from coking operations.

In 1988, California refineries had a combined rated charge capacity of approximately 2.28 million bbls per calendar day (Oil & Gas Journal (OGJ, 1989). According to OGJ, catalytic cracking units had a rated charge capacity of 645,500 bbls per stream day and thermal operations (assuming coking operations) had a rated charge capacity of 503,200 bbls per stream day in 1988 (OGJ, 1989). There are no data to estimate the exact amount of fresh feed going into either FCC or TCC units. However, EPA estimated that 94 percent of the fresh feed in 1979 is used in FCC units and that the TCC units have become obsolete since 1979 (EPA, 1987). To be conservative (because the acetaldehyde emission factor for FCC units is greater than that of for TCC units), the staff assumes that all of the fresh feed going to catalytic cracking is going into the FCC units.

Using data on feed rates to catalytic cracking units and to thermal operations and applying the appropriate emission factors, the staff estimates that refineries emitted approximately 190 to 770 tons of acetaldehyde in 1988 (see Appendix B).

c. Food Processing

In response to the Superfund Amendments and Reauthorization Act (SARA) of 1986, Section 313, a food processing facility in the San Francisco Bay Area Air Basin reported emissions of approximately 7 tons of acetaldehyde in 1990 (Rydbrink, 1991). This information has not been validated by the Bay Area Air Quality Management District (BAAQMD).

d. Coffee Roasting

In the coffee roasting process, green coffee beans are roasted to obtain the characteristic aroma and flavor associated with coffee. Aldehydes have been detected in green coffee beans; therefore aldehydes are emitted during the coffee roasting process. During roasting, green beans are exposed to temperatures as high as 410°F; chemical reactions such as pyrolysis or thermal decomposition occur within the beans releasing aldehydes such as acetaldehyde and other volatile organic compounds (MRI, 1987).

Based on the amount of green coffee beans roasted in California (ARB, 1990j) and a controlled emission factor for acetaldehyde developed from actual source tests by an Environmental Protection Agency (EPA) contractor

(MRI, 1987), the staff estimates that coffee bean roasters in California emitted approximately 2 tons of acetaldehyde in 1987 (see Appendix B).

5. Other Potential Sources

There are several potential sources of acetaldehyde in California which probably have very low emissions. Although the emission factors for these sources are unknown, they deserve some mention.

a. Coke Combustion

There are no acetaldehyde emission factors from coke combustion. However, based on the process rate and the acetaldehyde emission factor from coal combustion, the staff estimates that acetaldehyde emissions from coke combustion are insignificant.

b. Jet Fuel Combustion

Even though criteria pollutant emissions have been estimated for the combustion of jet fuel used at stationary IC engines and for aircraft, acetaldehyde emissions have not been measured. For aircraft, the districts estimated criteria pollutant emissions using the emission factors per landing and take off (LTO). The process rates in terms of the fuel used for several airfields are not available. However, using process rates for other sources in this category and an acetaldehyde emission factor from oil combustion, the staff estimates that acetaldehyde emissions from jet fuel combustion are insignificant.

c. Liquid Material Combustion

The ARB's EDS lists several stationary point sources burning unspecified liquid material. The staff believes that this category includes the combustion of diesel, distillate, jet fuel and residual oil. Based on the relative activities for these liquid material combustion sources, the staff estimates that of the 42.6 million gallons of liquid material burned, approximately 5.1 million gallons may be jet fuel combustion; the other 37.5 million gallons can be diesel, distillate, or residual oil (ARB, 1990i).

If acetaldehyde emission factors for oil are applicable to the liquid material combustion category, the staff estimates that acetaldehyde emissions from this category are not significant.

d. Liquified Petroleum Gas (LPG)

From the EDS, the staff has identified several point sources in California that burn LPG. The districts estimated criteria pollutant emissions for these combustion sources. The staff believes these combustion sources emit aldehydes and acetaldehyde in particular. However, based on the amount of LPG burned in California, the staff estimates that acetaldehyde emissions from these combustion sources are not significant.

e. Process Gas Combustion

Several stationary point sources in California reported emissions from the combustion of process gas. Because the constituents of each process gas are unique, the acetaldehyde emissions are unique for individual process gas combustion sources. Streams of process gas may even be different from one another within a facility. Thus emissions of process gas can not be generalized and acetaldehyde emissions for these combustion sources are not available at this time.

However, based on the amount of process gas burned at stationary sources in California, and assuming that the emission factors from natural gas combustion are applicable for these process gas combustion sources, the staff estimates that process gas combustion sources could emit approximately 50 percent of the estimated emissions from natural gas combustion.

f. Landfill Gas Combustion

Based on the amount of landfill gases burned in California, the staff estimates that acetaldehyde emissions from landfill gas combustion sources are not significant.

g. Incineration

Incineration has been considered an alternative method to dispose of waste and may be used as a method to recover energy to produce electricity. There are municipal waste, sewage sludge, hazardous waste and biomedical waste incinerators throughout the State. Some of these incinerators have been tested for dioxins and dibenzofurans emissions. However, there are no data on acetaldehyde emissions. Because the process rates of incinerators are relatively small as compared to fuel combustion processes such as coal combustion and because control measures are being proposed for incinerators, the staff estimates that acetaldehyde emissions from incinerators are relatively insignificant.

D. References for Chapter III

Air Resources Board (ARB) (1988) "Methodology to Calculate Emission Factors for On-road Motor Vehicles" - Draft Supplement 1. Emission Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1989) "Identification of Volatile Organic Compound Species Profiles - ARB Speciation Manual: Volume 1 of 2 ", Emission Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1990a) Data retrieved from the Emission Data System (EDS) EMFAC7E/BURDEN7C, AR7C1987(J3173) - Run date 3/27/90, Emission Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1990b) Data retrieved from the EDS regarding acetaldehyde emissions from area and mobile sources without emissions from activities occurring in the outer continental shelf for inventory year 1987. ARACEAMV(J6203) - Run date 7/11/90, Emission Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1990c) Data retrieved from the EDS regarding acetaldehyde emissions from area and mobile sources (for activities occurring in the outer continental shelf area only) for inventory year 1987. ARACEMV(J6091) - Run date 7/11/90, Emission Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1990d) Data retrieved from EMFAC7E/BURDEN7C for inventory year 1989, Run date: 8/9/90, Emissions Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1990e) Data retrieved from EMFAC7E/BURDEN7C for inventory 1987, Run date: 7/12/90, Emission Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1990f) Emission Inventory 1987 - March 1990, Emission Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1990g) Data retrieved from EDS, "1987 Area Source Emissions", Run date 6/11/90, Emission Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1990h) Data retrieved from the EDS regarding acetaldehyde emissions for inventory year 1987, ARACETST(J8558) - Run date 7/11/90, Emission Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1990i) Data retrieved from the EDS regarding fuel combustion, ARBAPF(J3265) - Run date 2/24/90, Emission Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1990j) Data retrieved from the EDS regarding coffee bean roasting in the State, ARACET(J2888) - Run date 3/24/90, Emission Inventory Branch, Technical Support Division, Sacramento, CA.

ARB (1991a) Emission Inventory 1989 - August 1991, Emission Inventory Branch, Technical Support Divison, Sacramento, CA.

ARB (1991b) Proposed Reactivity Adjustment Factors for Transitional Low-Emission Vehicles, Technical Support Document, Mobile Source Division, Research Division, Sacramento, CA.

ARB (1991c) Proposed Regulations for California Phase 2 Reformulated Gasoline, Technical Support Document, Stationary Source Division, Sacramento, CA.

Atkinson, R. (1993) Personal Communication

Atkinson, R. et al. (1990) "Lifetimes and Fates of Toxic Air Contaminants In California's Atmosphere", Final Report. Prepared for the ARB, Contract # A732-107, Statewide Air Pollution Research Center, University of California, Riverside, CA.

"Auto/Oil Air Quality Improvement Research Program", Technical Bulletins 1-5, 1990-1991.

Carey, P. (1981) "Mobile Source Emissions of Formaldehyde and Other Aldehydes", Control Technology Assessment and Characterization Branch, Emission Control Technology Division, Office of Mobile Source Air Pollution Control, United States Environmental Protection Agency (EPA), Ann Arbor, MI.

Environmental Protection Agency (EPA) (1980) "Assessment of Atmospheric Emissions From Petroleum Refining, Volume 1, Technical Report", EPA-600/2-80-075a, Research Triangle Park, NC.

EPA (1984) "Locating and Estimating Air Emissions from Sources of Formaldehyde", EPA-450/4-84-007e, Research Triangle Park, NC.

EPA (1987) "Locating and Estimating Air Emissions from Sources of Polycyclic Organic Matter (POM)", EPA-450/4-84-007p, Research Triangle Park, NC.

EPA (1988) "Toxic Air Pollutant Emission Factors - A Compilation for Selected Air Toxic Compounds and Sources", EPA-450/2-88-006a, Research Triangle Park, NC.

EPA (1992) "Motor Vehicle-Related Air Toxics Study", Public Review Draft, Technical Support Branch, Emission Planning and Strategies Division, Ann Arbor, MI.

EPA (1993) "Motor Vehicle-Related Air Toxics Study", Technical Support Branch, Emission Planning and Strategies Division, Ann Arbor, MI.

Finlayson-Pitts, B. J. and J. N. Pitts, Jr. (1986) "Atmospheric Chemistry: Fundamentals and Experimental Techniques", Published by John Wiley & Sons, Inc.

Grosjean, D., Miguel, A. and Tavares, T. (1990) "Urban Air Pollution in Brazil: Acetaldehyde and Other Carbonyls", Atmospheric Environment, 24B:1 pp. 101-106.

Ligocki, M. and G. Whitten (1991) "Atmospheric Transformation of Air Toxics: Acetaldehyde and Polycyclic Organic Matter", SYSAPP-91/113, Systems Applications International Corporation, San Rafael, CA 94903, prepared for USEPA, Office of Policy Planning and Evaluation, Washington, D.C.

Midwest Research Institute (MRI) (1987) "Source Assessment of Acetaldehyde Emissions", Draft Final Report, Research Triangle Park, NC.

Oil and Gas Journal (1989) "Annual Refining Survey", March 20, pp. 49-58.

Rydbrink, K. (1991) Personal communication between Linda Martz of the ARB and Ken Rydbrink of the Environmental Affairs Agency regarding acetaldehyde emissions as reported by California facilities according to the requirements of the Superfund Amendments and Reauthorization Act (SARA), Section 313, Feb. 6.

Science Applications International Corporation (SAIC) (1987) "Evaluation of Potential Toxic Air Contaminants - Phase I. Final Report", Manhattan Beach, CA.

Serth, R., Tierney, D. and Hughes, T. (1978) "Source Assessment: Acrylic Acid Manufacture - State of the Art", Monsanto Research Corporation for EPA, EPA-600/2-78-004W.

Sigsby, J., Tejada, S. and Ray, W. (1987) "Volatile Organic Compound Emissions from 46 In-Use Passenger Cars, USEPA, Environmental Sciences Research Laboratory, Research Triangle Park, NC.

SRI International (1989) "1989 Directory of Chemical Producers - United States", SRI International, Menlo Park, CA.

Springer, K. (1979) "Characterization of Sulfates, Odor, Smoke, POM and Particulates from Light and Heavy Duty Engines - Part IX", Southwest Research Institute, San Antonio, TX, for EPA, EPA-460/3-79-007.

Tanner, R. (1988) "Atmospheric Chemistry of Aldehydes: Enhanced Peroxyacetyl Nitrate Formation from Ethanol-Fueled Vehicular Emissions", Environmental Science and Technology, 22, pp. 1026-1034.

Ullman, T. and C. Hare (1985) "Emissions Characterization of Heavy-Duty Diesel and Gasoline Engines and Vehicles", Southwest Research Institute, San Antonio, TX, for EPA, EPA-460/3-85-001.

Urban, C. (1980a) "Regulated and Unregulated Exhaust Emissions from Malfunctioning Non-Catalyst and Oxidation Catalyst Gasoline Automobiles", Southwest Research Institute, San Antonio, TX, for EPA, EPA-460/3-80-003.

Urban, C. (1980b) "Regulated and Unregulated Exhaust Emissions from Malfunctioning Three-Way Catalyst Gasoline Automobiles", Southwest Research Institute, San Antonio, TX, for EPA, EPA-460/3-80-004.

Urban, C. (1980c) "Regulated and Unregulated Exhaust Emissions from a Malfunctioning Three-Way Catalyst Gasoline Automobile", Southwest Research Institute, San Antonio, TX, for EPA, EPA-460/3-80-005.

Urban, C. (1981) "Unregulated Exhaust Emissions from Non-Catalyst Baseline Cars Under Malfunction Conditions", Southwest Research Institute, San Antonio, TX, for EPA, EPA-460/3-81-020.

Wagner, K. (1991) Internal ARB memorandum from Dr. Kit Wagner to Beth Schwehr, subject: Estimates of Secondary Ambient Acetaldehyde, May 21, 1991, Control Strategy Modeling Section, Technical Support Division, ARB, Sacramento, CA.

IV.

EXPOSURE TO ACETALDEHYDE

A. AMBIENT MONITORING IN CALIFORNIA

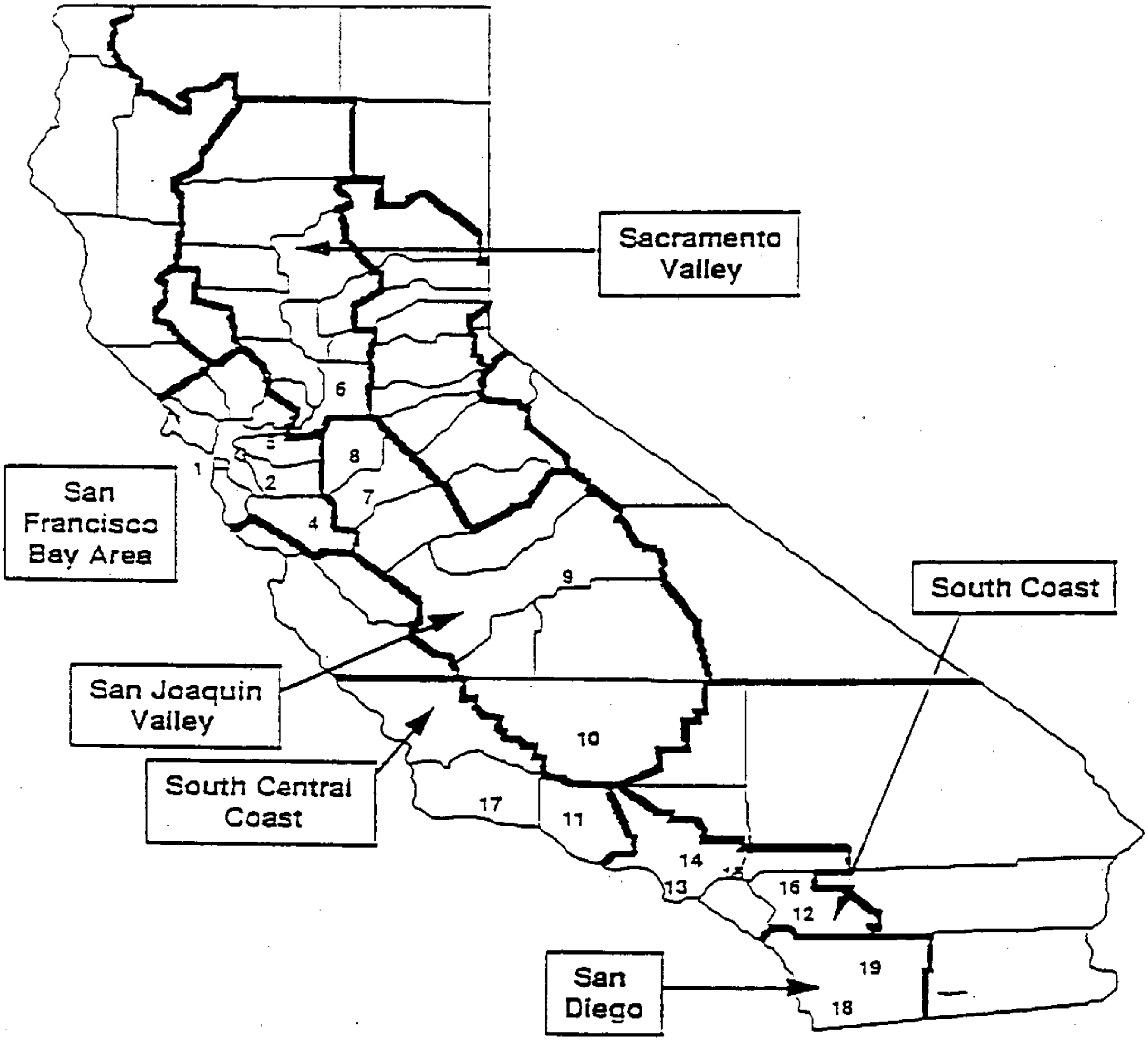
The toxics sampling network in California for acetaldehyde consisted of 19 monitoring stations statewide (Figure IV-1). Nine of these monitors were located in Southern California (south of Bakersfield), while the other 10 were situated in the northern portion of the state.

Data used in this exposure analysis were collected during the period of September 1988 through August 1989. The data for this period (hereafter referred to as "the study period") represent the most recent period for which the data are of consistent and verifiable quality. The data analysis and ambient exposure estimate are based on ambient data collected by the Air Resources Board (ARB) and maintained in the toxics air quality database.

Individual samples were collected over a 24 hour period using two cartridges containing 2,4-dinitrophenyl hydrazine coated support in tandem. The aldehydes present in the air react with the 2,4-dinitrophenylhydrazine coating to form the respective hydrazones. The hydrazone derivatives are extracted from the cartridges with acetonitrile, and the extract is analyzed using Reverse-Phase High Performance Liquid Chromatography with a ultraviolet detector. The limit of detection (LOD) of this procedure is 0.1 ppbv (0.18 $\mu g/m^3$). For the analytical procedure see Appendix C. Quality control is performed with all generated data sets and an extensive quality assurance program assures the accuracy of the measurements.

Figure IV-1

ARB Toxics Network Monitoring Sites



1. San Francisco
2. Fremont
3. Richmond
4. San Jose
5. Concord
6. Citrus Heights
7. Modesto
8. Stockton
9. Fresno
10. Bakersfield
11. Simi Valley
12. Riverside
13. Long Beach
14. Los Angeles
15. El Monte
16. Upland
17. Santa Barbara
18. Chula Vista
19. El Cajon

A summary of data for each of the monitoring sites used in the analysis is presented in Table IV-1. The statewide acetaldehyde data for the study period represent 19 sites and a small percentage of all possible days during the sampling period. The number of samples available per site during the study period range from 21 to 26 in northern California, averaging 24 observations per site for the study period; the number of samples available per site in southern California range from 21 to 57, averaging 40 observations per site.

Acetaldehyde was sampled on an every sixth day schedule in the South Coast Air Basin. This increase in sampling frequency accounts for the increased number of observations recorded for the South Coast in comparison to the other five air basins. There were no concentrations reported below the LOD.

B. AMBIENT CONCENTRATIONS

Summary statistics for each site during the study period are summarized in Table IV-2. The minimum, maximum, median, and estimated mean annual concentration, along with the coefficient of variation and standard deviation are reported for each site. Basin results are included for each statistic except the coefficient of variation.

Mean annual acetaldehyde concentrations were calculated as the mean of available monthly means. This approach provides equal weighting for each month even when the number of samples per month varies. The mean of monthly means is a more reliable estimator of annual exposure than the arithmetic average of all study period concentrations. Mean annual concentrations ranged from a minimum of 1.13 ppbv (2.03 $\mu g/m^3$) at Richmond to a maximum of 3.31 ppbv (5.96 $\mu g/m^3$) at El Monte. Basin mean concentrations ranged from 1.63 ppbv (2.93 μg/m3) in the South Central Coast to 2.81 ppbv (5.06 μg/m3) in the South Coast.

Table IV-1

SUMMARY OF ACETALDEHYDE DATA [1]
September 1988 - August 1989

Site Location	S	O	N	D	J	F	M	A	M	J	J	A	Samples
SOUTHERN CALIFORNIA SITES													
South Coast Air Basin													
El Monte	o	o	o	o	o	o	o	o	o	o			45
Long Beach	o	o	o	o	o	o	o	o	o	o	o	o	57
Los Angeles	o	o	o	o	o	o	o	o	o	o	o	o	56
Rubidoux	o	o	o	o	o	o	o	o	o	o	o	o	57
Upland	o	o	o	o	o	o	o	o	o	o	o	o	51
South Central Coast Air Basin													
Santa Barbara	o	o	o	o	o	o	o	o	o	o	o	o	21
Simi Valley	o	o	o	o	o	o	o	o	o	o	o	o	26
San Diego Air Basin													
Chula Vista	o	o	o	o	o	o	o	o	o	o	o	o	25
El Cajon	o	o	o	o	o	o	o	o	o	o	o	o	25
NORTHERN CALIFORNIA SITES													
San Francisco Bay Area Air Basin													
Concord	o	o	o	o	o	o	o	o	o	o	o	o	25
Fremont	o	o	o	o	o	o	o	o	o	o	o	o	25
Richmond	o	o	o	o	o	o	o	o	o	o	o	o	24
San Francisco		o	o	o	o	o	o	o	o	o	o	o	21
San Jose	o	o	o	o	o	o	o	o	o	o	o	o	23
San Joaquin Valley Air Basin													
Bakersfield	o	o	o	o	o	o	o	o	o	o	o	o	25
Fresno	o	o	o	o	o	o	o	o	o	o	o	o	26
Modesto	o	o	o	o	o	o	o	o	o	o	o	o	25
Stockton	o	o	o	o	o	o	o	o	o	o	o	o	24
Sacramento Valley Air Basin													
Citrus Heights	o	o	o	o	o	o	o	o	o	o	o	o	24

---

[1] A "o" indicates at least one sample was collected during the month.

Table IV-2

SUMMARY OF AMBIENT ACETALDEHYDE CONCENTRATIONS
September 1988 - August 1989
(parts per billion)

AIR BASIN Site Location	Minimum Conc.	Maximum Conc.	Median Conc.	Mean[1] Conc.	Std.[2] Dev.	CV[3]
SOUTHERN CALIFORNIA SITES						
South Coast Air Basin						
El Monte	0.9	8.8	2.8	3.31	1.03	52.2
Long Beach	0.5	7.7	2.1	2.49	1.28	69.7
Los Angeles	0.1	8.8	2.5	2.97	0.98	61.8
Rubidoux	0.3	6.6	2.3	2.40	0.94	54.8
Upland	0.2	7.7	2.7	2.88	0.63	46.6
Basin Summary	0.1	8.8	2.5	2.81	0.99	
South Central Coast Air Basin						
Santa Barbara	0.7	3.4	1.3	1.47	0.70	50.6
Simi Valley	0.3	4.0	1.9	1.79	0.72	58.1
Basin Summary	0.3	4.0	1.5	1.63	0.71	
San Diego Air Basin						
Chula Vista	0.6	4.5	1.0	1.39	0.78	64.7
El Cajon	0.8	6.1	2.3	2.49	0.91	53.5
Basin Summary	0.6	6.1	1.7	1.94	0.85	
NORTHERN CALIFORNIA SITES						
San Francisco Bay Area Air Basin						
Concord	0.3	3.9	1.4	1.77	0.87	55.0
Fremont	0.6	6.1	1.2	1.74	1.04	73.0
Richmond	0.4	2.3	1.0	1.13	0.41	46.4
San Francisco	0.3	7.7	1.2	1.95	1.30	93.8
San Jose	0.2	6.1	1.5	1.62	1.01	75.2
Basin Summary	0.2	7.7	1.3	1.64	0.97	
San Joaquin Valley Air Basin						
Bakersfield	0.2	3.4	1.3	1.45	0.58	62.6
Fresno	0.3	6.3	2.0	2.26	1.03	63.6
Modesto	0.3	7.2	1.6	2.00	1.11	73.7
Stockton	0.2	4.7	1.9	1.97	0.73	55.1
Basin Summary	0.2	7.2	1.6	1.98	0.89	
Sacramento Valley Air Basin						
Citrus Heights	0.5	4.1	2.1	2.01	0.88	50.2

---

1 Basin Means are the mean of the site means.

2 Basin Standard Deviations are pooled values of the standard deviations across sites within a basin.

3 Coefficient of Variation is calculated as the standard deviation of all values divided by the mean of all values expressed as a percentage. Basin results are not given.

The median is defined as the 50th percentile and can be used as an alternative measure of central tendency to the mean. The median was less than the mean at 18 of the 19 monitoring sites. Citrus Heights and Simi Valley were the only sites reporting a median higher than the mean. During the study period, median concentrations ranged from a minimum of 1.0 ppbv (1.8 $\mu g/m^3$) at Chula Vista and Richmond to a maximum of 2.8 ppbv (5.0 $\mu g/m^3$) at El Monte. Basin medians were calculated as the median of all values reported for each individual basin. The basin-wide median concentrations range from a minimum of 1.3 ppbv (2.34 $\mu g/m^3$) for the San Francisco Bay Area to a maximum of 2.5 ppbv (4.5 $\mu g/m^3$) for the South Coast. When the median concentration is significantly different from the mean this indicates the data most likely come from a non-normal distribution. The Shapiro-Wilk (Shapiro and Wilk, 1965) test was used on a site-by-site basis to test the distribution of the acetaldehyde data. Even though the data from most sites are skewed, the Shapiro-Wilk test showed that the acetaldehyde concentrations are not log-normally distributed.

A coefficient of variation (CV) statistic provides information about the distribution of data and is reported for each monitoring site. The CV expresses as a percentage the magnitude of the variation relative to the magnitude of the concentrations being measured. The CV is equal to the standard deviation of all observations divided by the mean of all observations multiplied by 100. The CV for the study period ranges from a low of 46.4 percent at Richmond to a high of 93.8 percent at San Francisco. There appears to be no significant difference in the range of values for the CV between northern and southern California indicating that the distribution of values is the same.

The observed distribution of values from each site is presented graphically in Figure IV-2. In general, the southern California sites appear to have slightly higher concentrations than the northern California sites. Figure IV-3 presents the same data on a month-by-month basis without regard to sites. Figure IV-3 suggests there is a seasonal pattern

FIGURE IV-2

DISTRIBUTION OF AMBIENT ACETALDEHYDE CONCENTRATIONS PLOTTED USING EXTENDED BOX PLOTS

(BASED ON SEPTEMBER 1988 – AUGUST 1989 DATA)

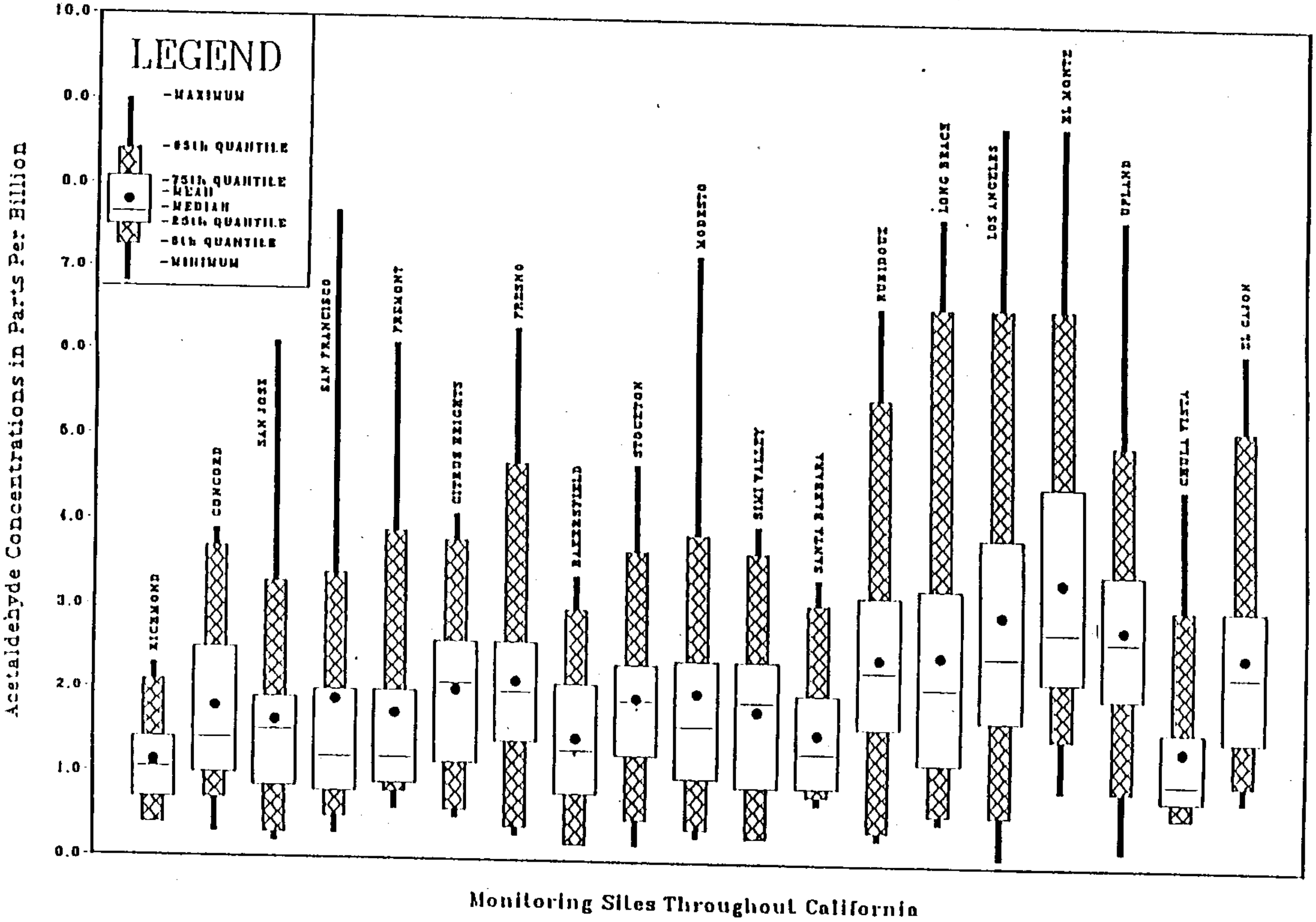
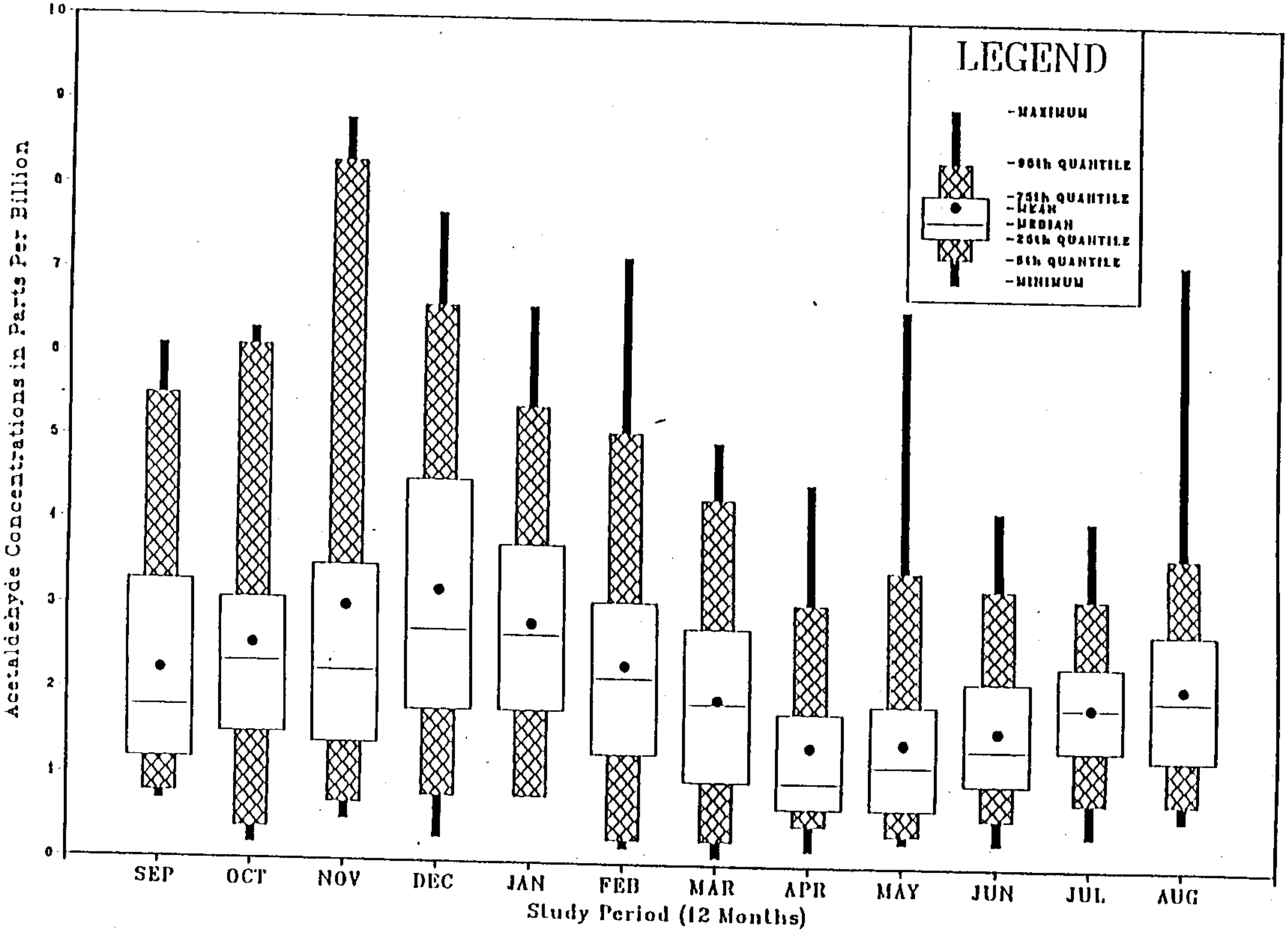


FIGURE IV-3

DISTRIBUTION OF MONTHLY AMBIENT ACETALDEHYDE CONCENTRATIONS PLOTTED USING EXTENDED BOX PLOTS (ALL SITES COMBINED)

BASED ON OCTOBER 1988 – SEPTEMBER 1989 DATA



ARB/TSD:April 90

to the distribution of ambient acetaldehyde concentrations. However, at this time we are unable to determine whether this pattern is the result of meteorological influences, changes in emission patterns, or fluctuations in secondary formation potential.

It may be expected that the ambient atmosphere acetaldehyde levels encountered in the Los Angeles air basin are among the highest to be expected in California because of the number of combustion sources associated with periods of high photochemical reactivity and the meteorological conditions present in the atmosphere. Reported data for the ambient concentrations of acetaldehyde in California since 1980 are given in Table IV-3. These figures demonstrate high smog events, varied hours of sampling times and variability between cities. Four studies performed by Grosjean revealed similiar results. These studies showed how acetaldehyde tends to have high levels in the daytime hours and low levels during the nighttime hours. During the Nitrogen Methods Comparison Study (NSMCS) At Claremont, CA, high levels of acetaldehyde existed during daytime hours and high motor vehicle travel. There were low levels at night (Grosjean, 1988). Peak concentrations ocurred at midday with concentrations of 6.1 ppbv (11 $\mu g/m^3$). The low occurred from midnight until dawn with concentrations of 2.1 ppbv (3.8 $\mu g/m^3$).

In 1987, acetaldehyde was measured during the South Coast Air Quality Study (SCAQS) in the South Coast Air Basin. The SCAQS was an integrated air quality study whose overall goal was to develop a comprehensive and properly archived air quality and meteorological data base for the South Coast Air Basin which can be used to test, evaluate, and improve elements of air quality simulation models for oxidants, NO_2, PM_{10}, fine particles, visibility, toxic air contaminants, and acidic species. Acetaldehyde concentrations ranged from 0.9 to 24.5 ppbv (1.6 to 44.1 $\mu g/m^3$) in 1 hour samples. In the South Coast Air Basin, on average, the levels peaked around midday at locations like Anaheim, Burbank, downtown Los Angeles, Hawthorne and Long Beach while they peaked later in the afternoon downwind at smog receptor sites like Azuza, Claremont, Rubidoux, Palm Springs, and Perris (Fung, 1989; Grosjean, 1991, 1992).

Table IV-3

MEASUREMENTS OF ACETALDEHYDE CONCENTRATIONS IN AMBIENT AIR IN CALIFORNIA

Acetaldehyde (ppbv)	Sampling Time (hours)	Measurement Date and Place	Reference
up to 26	0.5	5-6/1980, East Los Angeles	a
3-5	1	9-10/1980, Claremount	a
up to 10	1-2	10/1980, Azusa	b
up to 3	1-2	10/1980, Lennox	b
13-37	0.5	7-10/1980, Los Angeles	b
8-16	0.5	7-10/1980, Burbank	b
13	0.5	7/1980, Pasadena	b
25-33	0.5	10/1980, El Monte	b
16	0.5	10/1980, Rosemead	b
19	0.5	10/1980, Upland	b
2-39	0.45	9-11/1981, Los Angeles	c
2-15	2.0	2/1986, Downey	d
5.0	24.0	6/89-6/90, Palm Springs	e
10.3	24.0	6/89-6/90, Perris	e

---

a Grosjean, 1982.

b Grosjean, 1983.

c Grosjean and Fung, 1984.

d Salas and Singh, 1986

e Grosjean and Williams, 1992.

Reports of ambient air samples from two rural regions, Point Barrow, Alaska (Cavanagh et. al, 1969) and Whiteface Mountain, New York (Schulam et. al, 1984) provide examples of 'background' concentrations of 0.0 to 0.8 ppbv (1.4 $\mu g/m^3$) of acetaldehyde.

C. POPULATION WEIGHTED EXPOSURE ESTIMATES

Mean population exposure estimates were calculated using the study period acetaldehyde data. Exposure for the South Coast Air Basin and San Francisco Bay Area Air Basin were estimated by interpolating station values to census tract centroids. For the other air basins, a basin-wide mean concentration was estimated from the means for all sites in the basin.

It was then assumed that all people in those basins (except the Sacramento Valley Air Basin since it had only one monitoring station) with a sampling site were exposed to this estimated basin-wide mean concentration. Population data used in the exposure analysis represent 1980 census data updated to 1985. The results of the exposure analysis are summarized in Table IV-4. The table shows outdoor ambient annual exposures for different air basins and an overall population weighted exposure for the state of California.

Bootstrap confidence bounds are provided as a measure of uncertainty in the exposure estimates. The larger the difference between the mean annual exposure estimate and the upper and lower bounds, the greater the uncertainty in the estimate. The bootstrap method allows data contained in a single sample to be applied when estimating the accuracy of a statistical measure provides freedom from the assumption that the data conform to a predetermined distribution (Peterson, 1991) (Appendix D). Table IV-5 provides the bootstrap confidence bounds. These bounds represent the uncertainty in exposure estimates, not actual exposure. The upper and lower bounds used in this study were obtained through a random resampling of the monthly means at each site. The use of monthly means maintains compatability with the mean annual exposure estimates.

Table IV-4

SUMMARY OF MEAN ANNUAL ACETALDEHYDE EXPOSURE ESTIMATES
(parts per billion)

AIR BASIN	ESTIMATED EXPOSURE	POPULATION
SOUTHERN CALIFORNIA BASINS		
South Coast	2.85	10,092,133
South Central Coast	1.63	934,800
San Diego	1.94	2,131,600
NORTHERN CALIFORNIA BASINS		
San Francisco Bay Area	1.69	4,394,374
San Joaquin Valley	1.92	1,778,400
Sacramento Valley[1]	2.01	893,800
OVERALL POPULATION-WEIGHTED EXPOSURE		
	2.33	20,225,107

---

[1] Exposure estimates are for Sacramento County residents only.

Table IV-5

SUMMARY OF MEAN ANNUAL ACETALDEHYDE CONCENTRATIONS AND BOOTSTRAP ESTIMATES OF UNCERTAINTY
(parts per billion)

Air Basin Site Location	LOWER BOUND	ANNUAL MEAN	UPPER BOUND
SOUTHERN CALIFORNIA SITES			
South Coast Air Basin			
El Monte	2.73	3.31	3.84
Long Beach	1.78	2.49	3.14
Los Angeles	2.44	2.97	3.40
Rubidoux	1.86	2.40	2.84
Upland	2.55	2.88	3.19
South Central Coast Air Basin			
Santa Barbara	1.10	1.47	1.83
Simi Valley	1.42	1.79	2.17
San Diego Air Basin			
Chula Vista	0.93	1.39	1.78
El Cajon	1.99	2.49	2.93
NORTHERN CALIFORNIA SITES			
San Francisco Bay Area Air Basin			
Concord	1.30	1.77	2.21
Fremont	1.18	1.74	2.30
Richmond	0.90	1.13	1.33
San Francisco	1.11	1.95	2.55
San Jose	1.03	1.62	2.08
San Joaquin Valley Air Basin			
Bakersfield	1.11	1.45	1.72
Fresno	1.69	2.26	2.79
Modesto	1.35	2.00	2.57
Stockton	1.56	1.97	2.34
Sacramento Valley Air Basin [1]			
Citrus Heights	1.50	2.01	2.42

---

[1] Exposure estimates are for Sacramento County residents only.

The overall ambient outside statewide acetaldehyde exposure, weighted by population, is best estimated to be 2.33 ppbv (4.19 $\mu g/m^3$). A total of approximately 20 million people are estimated to reside in the study areas, representing approximately 80 percent of the State's population. Basin-specific, population-weighted mean concentrations vary from a minimum of 1.63 ppbv (2.93 $\mu g/m^3$) in the South Central Coast Air Basin to a maximum of 2.85 ppbv (5.13 $\mu g/m^3$) in the South Coast Air Basin (See Table IV-4). The South Coast also accounted for five of the six sites with the highest mean annual exposure estimates in California. Figure IV-4 shows the total number of people exposed to various mean annual acetaldehyde concentrations.

The geographic mean-annual acetaldehyde concentration calculated as the average of the six basin averages was 2.00 ppbv (3.6 $\mu g/m^3$). This value is approximately 15 percent lower than the population-weighted exposure best estimate of 2.33 ppbv (4.19 $\mu g/m^3$) indicating that the highest amounts of acetaldehyde tended to be in areas of higher population density.

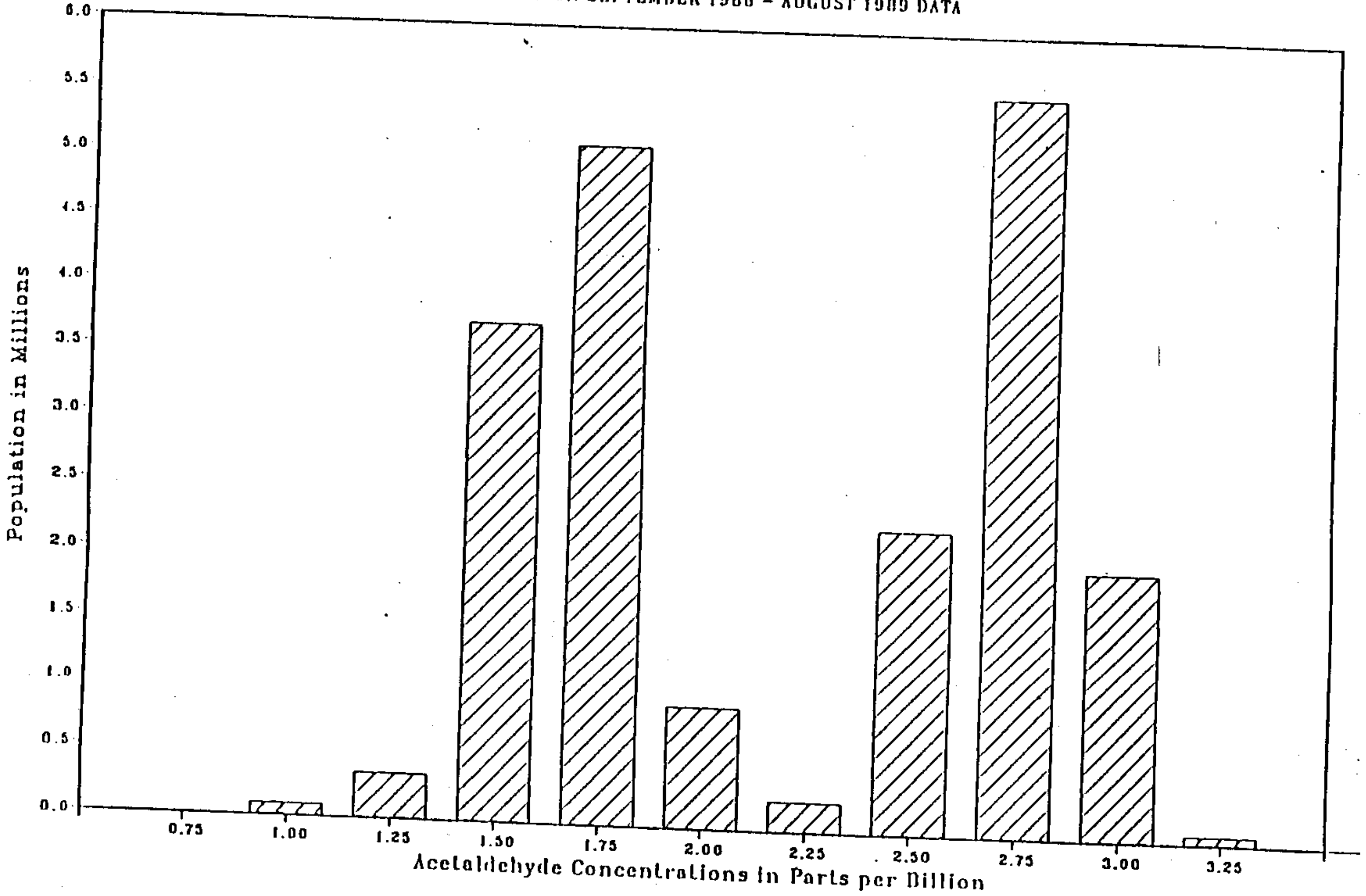
D. EXPOSURE TO ACETALDEHYDE NEAR EMISSION SOURCES

Some Californians may be exposed to near-source, or "Hot Spot" concentrations of acetaldehyde which are above the average ambient concentrations. "Hot Spot" exposure may increase the potential cancer risk to individuals living near large combustion sources. Acetaldehyde emissions and risk information is currently under development by facilities under AB 2588 Air Toxics "Hot Spots" emissions reporting program. This information will be used during the risk management phase to help determine priority and need for control of sources emitting acetaldehyde.

E. INDOOR EXPOSURE

Indoor exposure assessment has become increasingly important as an integral part of assessing exposure to toxic air contaminants because:

FIGURE IV-4
POPULATION ESTIMATES
FOR VARIOUS MEAN ANNUAL AMBIENT CONCENTRATIONS
BASED ON SEPTEMBER 1988 - AUGUST 1989 DATA



ARB/TSD:March 90

1. Californians spend most of their time (about 87 percent on average) indoors (Wiley et al., 1989); and

2. personal and indoor air monitoring data indicate that some pollutant concentrations are regularly higher indoors than outdoors.

Thus, consideration of indoor air exposure data is critical in developing a realistic estimate of personal exposures through the air environment.

1. Indoor Concentrations

Only a small number of limited surveys that measured acetaldehyde concentrations in indoor air have been published (summarized in Table IV-6). Those surveys monitored a small number of homes, and the homes that were monitored were not selected randomly. With such limitations, it is not possible to extrapolate the monitoring results to apply to the general population of homes in the respective areas surveyed. However, when viewed together, the results of individual surveys can at least provide a general idea of the magnitude of indoor acetaldehyde concentrations. A very crude estimate of an average acetaldehyde concentration that would be expected to occur inside residences is about 3 to 15 ppbv (~ 5 to 27 $\mu g/m^3$). The highest acetaldehyde concentration reported inside a residence was 37 ppbv (67 $\mu g/m^3$). The results of one study suggest that higher levels may occur in some indoor environments, in this case a tavern occupied by a number of people who were smoking, where levels of up to 113 ppbv (203 $\mu g/m^3$) were reported. Average and maximum in-vehicle acetaldehyde concentrations [~ 8 and 37 ppbv, (~ 14 and 67 $\mu g/m^3$) respectively] appear to be similar in magnitude to those inside residences (Shikiya et al., 1989). Extremely limited data suggest that acetaldehyde concentrations in offices and public buildings are also similar in magnitude to those inside residences (Bufalini, 1990; Druzik et al., 1990).

Table IV-6

INDOOR AIR CONCENTRATIONS OF ACETALDEHYDE*
(parts per billion)

LOCATION	AVERAGE CONCENTRATION	RANGE	N	COMMENTS	REFERENCES
<u>RESIDENCES</u>					
Boise, ID	9.3-11.8		20	non-smokers	Zweidinger et al., 1988
California		2.6 & 7.1	2		Rogozen et al., 1984
Raleigh, NC	4.0-8.5		3	non-smokers	Zweidinger, et al., 1987 Highsmith, et al., 1988
Roanoke, VA	8.3	3-26**	20	non-smokers	Zweidinger, personal communication
Baltimore, MD	14.8	6.2-37$	8	smokers may be included	Nelson, personal communication
Raleigh/ Durham, NC	20.4 15.2		15 12	+3 mobile -3 mobile	Tejada, personal communication
Raleigh/ Durham, NC		3.6-15	2		Bufalini, personal communication
<u>PUBLIC BLDGS.</u>					
Los Angeles, CA 5 museums, 1 library		4.2-35	6		Druzik et al., 1990
North Carolina office bldg.		1.2 12.8	1 1	non-smoking smoking	Bufalini, personal communication
North Carolina tavern		102 & 113	1	heavy smoking	Lofroth et al., 1989

(continued on next page)

INDOOR AIR CONCENTRATIONS OF ACETALDEHYDE continued

LOCATION	AVERAGE CONCENTRATION	RANGE	N	COMMENTS	REFERENCES
INSIDE VEHICLES					
Southern California	7.6	max. 37.0	194	commute driving	Shikiya et al., 1989
SMOKING CHAMBER					
		48 & 52		1 cig./30 min.	Lofroth et al., 1989
		114 & 129		1 cig./15 min.	

---

* Data on indoor air concentrations are limited; data cannot be extrapolated to apply to general population of homes in the respective areas surveyed. The data presented here are intended to provide a general idea of the magnitude of indoor acetaldehyde concentrations.

** Contamination of some field blanks casts doubt on the validity of some of the high values.

2. Indoor Sources

Acetaldehyde is formed as a combustion by-product and can be emitted from a number of indoor sources including cigarettes, fireplaces, woodstoves, and cooking. Limited information indicates that acetaldehyde also is emitted from some building materials such as rigid polyurethane foam insulation (Hodgson and Wooley, 1991). It may be present in a variety of consumer products such as adhesives, coatings, lubricants, inks, and nail polish remover (ibid.). A concentration of 1060 ppm has been detected in vinegar (Feron et al., 1991). Acetaldehyde occurs in traces in all ripe fruit and may form in alcoholic beverages after exposure to air (Fishbein, 1979), although it is not known if volatilization from those sources could affect indoor acetaldehyde concentrations to any significant degree.

3. Ingestion

Acetaldehyde is found commonly in a number of foods as a natural constituent and as an intentional food additive. The daily amount ingested through food cannot be estimated using readily available information. Levels in water are relatively low; a high estimate of the amount a Californian may ingest through water is 8 μg/day. See Appendix E. Dermal absorption is surmised to be negligible. It should be noted that, with respect to the potential carcinogenicity of acetaldehyde, the ingestion route may not be of major significance. There are well-developed means of detoxifying absorbed acetaldehyde; it is surmised that only in specialized organs such as the nasal cavity that inhalation of appreciable amounts may overcome local defense mechanisms (Casarett et al., 1986).

4. Inhalation

The inhalation route is of major importance to the toxic (e.g. carcinogenic) effects of acetaldehyde. Estimates of the daily amounts of acetaldehyde inhaled through residential indoor air and from environmental tobacco smoke are shown in Table IV-7. These estimates are based on very limited data and are thus very crude. No attempt has been made to use any estimates of the fraction absorbed so that the already high level of uncertainty in the values reported is not increased further. See Appendix E.

Table IV-7

ESTIMATES OF THE INHALED DOSE OF ACETALDEHYDE FROM RESIDENTIAL INDOOR AIR AND ENVIRONMENTAL TOBACCO SMOKE

	Acetaldehyde (micrograms per day)
Residential indoor air*	
a. average amount inhaled	108 - 540
b. high range of estimate	1,440
Environmental tobacco smoke**	1,740 - 4,660

* Based on very limited data; office building and in-vehicle exposures are estimated to be very similar to residential exposures.

** Based on chamber and tavern concentrations measured by Lofroth et al. (1989) under specific conditions; does not necessarily represent range of all possible levels.

Assumptions:

1. The average person inhales about 20 cubic meters of air per day.
2. Ingestion and skin absorption are not significant routes of exposure.
3. A person inhales about the same acetaldehyde concentration all day.
4. The inhaled dose may not represent the true dose, since it does not take into account the fraction of the inhaled amount that is absorbed.

5. Contribution to Total Exposures

Acetaldehyde concentrations are generally higher indoors than outdoors. This is likely due to a variety of indoor sources of acetaldehyde. Table IV-8 shows that indoor acetaldehyde concentrations measured in recent surveys range from 2.6 to 37.4 ppbv (4.7 to 67.3 $\mu g/m^3$) inside residences. In contrast with available indoor data, ARB ambient monitoring data (Table IV-2) show that the highest 24 hour average ambient outdoor acetaldehyde concentration measured was 8.8 ppbv (15.8 $\mu g/m^3$) recorded at the Los Angeles and El Monte stations. Although some of the difference between these indoor and outdoor data may be due to the differences in measurement techniques and other factors, indoor concentrations have been consistently higher in surveys that obtained concurrent indoor and outdoor acetaldehyde measurements. Table IV-8 compares indoor and outdoor acetaldehyde concentrations measured concurrently in some of those surveys. Although the data presented are derived from limited surveys, they show that indoor concentrations can be about two to eight times higher than outdoor concentrations. The sites surveyed may not exhibit acetaldehyde levels representative of California residences and offices because of different building construction, heating practices, energy consumption and inhabitants' lifestyles.

6. Summary

Individuals are estimated to inhale 108-540 μg of acetaldehyde daily in indoor environments where inhaled doses may range up to 1,440 μg. If a person were to spend the entire day in a smoke-filled environment such as a tavern, the daily inhaled dose could total from 1,740 to 4,660 μg. These estimates are all based on very limited data and may not be representative of the California population. A more detailed review of indoor acetaldehyde is included in Appendix E.

Table IV-8

CONCURRENT INDOOR AND OUTDOOR ACETALDEHYDE CONCENTRATIONS

Range of indoor concentrations (ppbv) 12 hour sample	Outdoor concentrations (ppbv) 12 hour sample	Reference
9.3 - 11.8	2.1 - 2.6 $	Zweidinger et al., 1988 Boise, ID
4.0 - 8.5	1.3 - 4.4	Zweidinger et al., 1987 Raleigh, NC
6.2 - 37.4	0.3 - 4.1	Nelson, personal communication, 1990 Baltimore, MD

F. References for Chapter IV

Bufalini, J. (1990) Environmental Protection Agency, Atmospheric Sciences Research Laboratory, North Carolina, personal communication.

California State Board of Equalization, Department of Research, Cigarette Taxes, June, 1990.

Cavanagh, L., Schadt, C. and Robinson, E. (1969) "Atmospheric Hydrocarbon and Carbon Monoxide Measurements at Point Barrow, Alaska," Environmental Science and Technology, 3: 251-257.

Casarett, L., Doull, J., Klaassen, C. and Amdur, M. eds. (1986) Casasrett and Doull's Toxicology: The Basic Science of Poisons, third edition, Macmillan Publishing Company, New York, p. 776.

Druzik, C., Grosjean, D., Van Neste, A. and Parmer, S. (1990) "Sampling of Atmospheric Carbonyls with Small DNPH-Coated C18 Cartridges and Liquid Chromatography Analysis with Diode Array Detection," International Journal of Environ. Anal. Chem., 38: 495-512.

Efron, B. and Tibshirani, R. (1986) "Bootstrap Methods for Standard Errors, Confidence Intervals and Other Measures of Statistical Accuracy," Statistical Science, 1(1):54-77.

Feron, V., Til, H., Flora de Vrijer, Woutersen, R., Cassee, F. and van Bladeren, P. (1991) "Aldehydes: Occurrence, Carcinogenic Potential, Mechanism of Action and Risk Assessment," Mutation Research, 259:363-385.

Fishbein, L. (1979) Potential Industrial Carcinogens and Mutagens, Elsevier Scientific Publishing Company, New York, pp. 149-150.

Fung, K., (1989) "Carbonyl Observations During the SQAQS," Presesntation at the 82nd Annual Meeting & Exhibition of the Air & Waste Management Association, Anaheim, CA, June, 25.

Grosjean, D. (1982) "Formaldehyde and Other Carbonyls in Los Angeles Ambient Air," Environmental Science and Technology, 16:254-262.

Grosjean, D., Swanson, R. and Ellis, C. (1983) "Carbonyls in Los Angeles Air:Contribution of Direct Emissions and Photochemistry," Science of the Total Environment, 28:65-85.

Grosjean, D. and Fung, K. (1984) "Hydrocarbons and Carbonyls in Los Angeles Air," Journal of the Air Pollution Control Association, 34:537-543.

Grosjean, D. (1988) "Aldehydes, Carboxylic Acids and Inorganic Nitrate During NSMCS," Atmospheric Environment, 22(8), pp. 1637-1648.

Grosjean, D. (1991) "Ambient Levels of Formaldehyde, Acetaldehyde, and Formic Acid in Southern California: Results of a One-Year Base-Line Study," Environmental Science and Technology, 25:710-715.

Grosjean, D. and Williams II, E. L. (1992) "Photochemical Pollution at Two Southern California Smog Receptor Sites", Journal of the Air & Waste Management Association, Vol. 42, pp. 805-809.

Highsmith, V. and Zweidinger, R. (1988) "Characterization of Indoor and Outdoor Air Associated with Residences Using Woodstoves: A Pilot Study," Environmental International, 14:213-219.

Hodgson, A. and Wooley, J. (1991) "Assessment of Indoor Concentrations, Indoor Sources and Source Emissions of Selected Volatile Organic Compounds," Final Report, Air Resources Board Contract No. A933-063, pp 63-69.

Lipari, F., Dasch, J. and Scruggs, W. (1984) "Aldehyde Emissions from Wood-Burning Fireplaces," Environmental Science and Technology, 18(5): 326-329.

Lofroth, G., Burton, R., Forehand, L., Hammond, S., Sella, R., Zweidinger, R. and Lewtas, J. (1989) "Characterization of Environmental Tobacco Smoke," Environmental Science and Technology, 23(5):610-614.

Nelson, W. (1990) Environmental Protection Agency, Environmental Monitoring Systems Laboratory, personal communication.

Peterson, I. (1991) "Pick a Sample", Science News, 140(4):56-58.

Rogozen, M., Maldonado, G., Grosjean, D., Shoshet, A. and Rappaport, R. (1984) "Formaldehyde: A Survey of Airborne Concentrations and Sources," Final report to ARB, contract no. A2-059-32.

Salas, L. and Singh, H. (1986) "Measurements of Formaldehyde and Acetaldehyde in the Urban Ambient Air", Short Communications, Atmospheric Environment, 20:6, pp. 1301-1304.

Schulam, P., Newbold, R. and Hull, L. (1985) "Urban and Rural Ambient Air Aldehyde Levels in Schenectady, New York and on Whiteface Mountain, New York," Atmospheric Environment, 19:4, pp. 623-626.

Shapiro, S. and Wilk, M. (1965) "An Analysis of Variance Test for Normality (Complete Samples)," Biometrica, 52:591.

Shikiya, D., Chung, S., Kahn, M., Juarros, J. and Barcikowski, W. (1989) "In-Vehicle Air Toxics Characterization Study in the South Coast Air Basin," South Coast Air Quality Management District.

Tejada, S. (1990) Environmental Protection Agency, Atmospheric Sciences Research Laboratory, personal communication.

United States Department of Health and Human Services (1989), Reducing the Health Consequences of Smoking: 25 Years of Progress: A Report of the Surgeon General.

World Health Organization (1985) IARC (International Agency for Research on Cancer) Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Man, Geneva, Vol. 36, pp.103-107.

Wiley, J., Robinson, J., Piazza, T., Garrett, K., Cirksena, K., Cheng, Y. and Martin, G. (1989) "Activity Patterns of California Residents," Final Report to ARB, Contract No. A6-177-33.

Zweidinger, R., Dropkin, D., Stump, F. and Tejada, S. (1987) "Volatile Organic Hydrocarbon and Aldehyde Composition in Raleigh, N.C. During The 1985 Woodsmoke Study," Proceedings of the 1987 EPA/APCA Symposium on the Measurement of Toxic and Related Air Pollutants, pp. 579-584.

Zweidinger, R., Tejada, S., Highsmith, R., Westburg, H. and Gage, L. (1988) "Distribution of Volatile Organic Hydrocarbons and Aldehydes During the IACP Boise, Idaho Residential Study," Proceedings of the 1988 EPA/APCA International Symposium on the Measurement of Toxic and Related Air Pollutants, pp. 814-820.

Zweidinger, R. (1990) Environmental Protection Agency, Atmospheric Sciences Research Laboratory, personal communication.

V.

ATMOSPHERIC FORMATION, PERSISTENCE, and FATE OF ACETALDEHYDE

The concentration of acetaldehyde in the atmosphere depends upon direct emissions, secondary formation, and chemical and physical removal activities. Direct emissions from mobile and stationary sources are discussed in Chapter IV while secondary formation, persistence, and fate of acetaldehyde are discussed here. The following summarizes the key points which are described in more detail in Sections A, B, and C:

- In polluted atmospheres, secondary acetaldehyde formation from the degradation of organic pollutants frequently dominates direct emissions by contributing 41 to 67 per cent of the total atmospheric acetaldehyde.

- Regardless of whether acetaldehyde is directly emitted or derived from secondary formation, its atmospheric persistence is generally one day which is sufficient time to allow dispersal throughout an air basin under normal meteorological conditions. However, it may take longer under multi-day stagnation episodes such as can occur in the South Coast Air Basin.

- The dominant atmospheric removal mechanism for acetaldehyde is oxidation by hydroxyl radicals during daylight hours.

- The end product of the acetaldehyde-hydroxyl radical reaction is another serious pollutant, peroxyacetyl nitrate (PAN).

A. ATMOSPHERIC FORMATION

Acetaldehyde formation from the degradation of organic pollutants in urban atmospheres contributes 41 to 67 per cent of the total atmospheric acetaldehyde. Therefore, secondary acetaldehyde formation frequently exceeds direct emissions from combustion sources in urban areas (Grosjean et al. 1983). In addition, the degradation of naturally-occurring ethane contributes a small amount of acetaldehyde to polluted and non-polluted atmospheres.

1. Acetaldehyde Formation in Polluted Atmospheres

In polluted atmospheres, major sources of ethyl radicals include the reaction of propionaldehyde with hydroxyl radicals

$$OH + CH_3CH_2CHO \rightarrow H_2O + CH_3CH_2CO$$

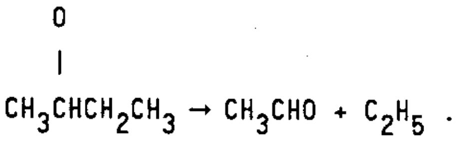
$$\downarrow O_2$$

$$CH_3CH_2C(O)OO$$

$$\downarrow NO$$

$$C_2H_5 + CO_2 ,$$

and the decomposition reactions of the more complex alkoxy radicals, e.g., 2-butoxy radical decomposition yields acetaldehyde as well as ethyl radicals



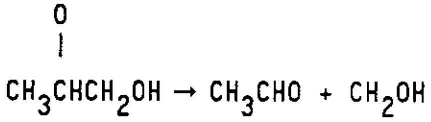
$$CH_3\overset{\overset{\displaystyle O}{|}}{C}HCH_2CH_3 \rightarrow CH_3CHO + C_2H_5 .$$

In each case the ethyl radical is oxidized to form acetaldehyde

$$C_2H_5 \rightarrow\rightarrow CH_3CH_2O$$

$$CH_3CH_2O + O_2 \rightarrow CH_3CHO + HO_2 \ .$$

Hydroxyl radical addition to propene and certain 2-alkenes forms β-hydroxyalkoxy radicals which decompose to form acetaldehyde or the α-hydroxy radical CH_3CHOH



$$CH_3\overset{O}{\overset{|}{C}}HCH_2OH \rightarrow CH_3CHO + CH_2OH$$

$$CH_3CHOHCH_2O \rightarrow CH_3CHOH + HCHO$$

The α-hydroxy radical is then oxidized to form acetaldehyde (Atkinson and Lloyd 1984, Atkinson 1989)

$$CH_3CHOH + O_2 \rightarrow CH_3CHO + HO_2 \ .$$

In addition, the gas-phase reactions of ozone with propene ($CH_3CH{=}CH_2$) and certain 2-alkenes leads to the formation of acetaldehyde

$$O_3 + CH_3CH{=}CH_2 \rightarrow CH_3CHO + HCHO + X$$

where X represents H_2O, OH, HO_2, H_2 and various hydrocarbons formed in much smaller amounts than acetaldehyde and formaldehyde (Atkinson 1990).

2. Acetaldehyde Formation in Non-polluted Atmospheres

In a non-polluted or "clean" troposphere, acetaldehyde is formed as a product of atmospheric ethane reactions which rapidly yield the ethyl peroxy radical ($CH_3CH_2O_2$). The ethyl peroxy radical reacts with nitric oxide, nitrogen dioxide, hydroperoxyl, and organic peroxy radicals leading to the formation of the ethoxy radical (CH_3OH_2O) which reacts with oxygen to form acetaldehyde

$$OH + C_2H_6 \rightarrow H_2O + C_2H_5$$

$$C_2H_5 \rightarrow\rightarrow CH_3CH_2O$$

$$CH_3CH_2O + O_2 \rightarrow CH_3CHO + HO_2$$

(Atkinson and Lloyd 1984, Atkinson 1990).

B. PERSISTENCE

The persistence of atmospheric acetaldehyde (the length of time acetaldehyde remains unaltered in the atmosphere) is determined by its tropospheric lifetime (τ). Tropospheric lifetime is the time required to decrease the concentration of acetaldehyde to 1/e or 0.368 of its initial concentration. The chemical reaction of acetaldehyde and hydroxyl radicals in the atmosphere is primarily responsible for the estimated one day tropospheric lifetime of acetaldehyde. Acetaldehyde removal by photolysis and physical means such as precipitation is expected to be less significant (Atkinson 1989b). All of these processes occur in the troposphere, the region of the atmosphere extending from the earth's surface to an altitude of approximately 15 km (Finlayson-Pitts and Pitts 1986).

1. Chemical Removal from the Atmosphere

The chemical removal of atmospheric acetaldehyde is the result of the sum of several distinct oxidative reactions and photolysis.

Acetaldehyde is oxidized in reactions with hydroxyl radicals, ozone, nitrate radicals, and hydroperoxyl radicals. Each of these chemical degradations of acetaldehyde is associated with a particular reaction speed determined by the reaction rate constant (k) and the atmospheric concentrations of reactants:

$$CH_3CHO + B \xrightarrow{k} \text{products}$$

where B is the atmospheric oxidant, and

$$\tau = 1/k[B].$$

Table V-1 lists the estimated average atmospheric lifetime of acetaldehyde relative to each oxidative removal reaction. Since the shortest lifetime is associated with the acetaldehyde-hydroxyl radical reaction, it is the single most important chemical removal mechanism. The acetaldehyde-hydroxyl radical reaction, photolysis, and the remaining oxidative reactions are discussed in order of importance in Subsections 1.a., 1.b., and 1.c., respectively.

a. Reaction with the Hydroxyl (OH) Radical

The photooxidation of acetaldehyde by hydroxyl radicals during daylight hours is considered the dominant atmospheric acetaldehyde removal mechanism because it produces the shortest lifetime for acetaldehyde

Table V-1

ATMOSPHERIC LIFETIME OF ACETALDEHYDE AND REMOVAL PROCESSES at 298°K

Oxidant	Atmospheric Concentration (mol cm^{-3})	Rate Constant ($cm^3 mol^{-1} sec^{-1}$)	Atmospheric Lifetime	References
Hydroxyl radical	1.5×10^6	1.58×10^{-11}	1 day[a]	Atkinson and Pitts 1978 Michael 1985 Niki et al. 1978 Prinn et al. 1987
Ozone	7×10^{11}	$< 1 \times 10^{-20}$	>4.5 yr.	Atkinson and Carter 1984 Logan 1985
Nitrate radical	2.4×10^8	2.7×10^{-15}	35 days	Dlugokencky and Howard 1989 Platt et al. 1984
Hydroperoxyl radical	$\sim 10^7$	1×10^{-15}	~3 yr.	Hard et al. 1984 Moortgat et al. 1987

a. Note that one day is equivalent to 12 hours of daylight. Hydroxyl reactions occur only during daylight hours.

(approximately one day). The short lifetime is due to the acetaldehyde-hydroxyl reaction rate constant and the concentration of hydroxyl radicals in the atmosphere

$$\tau = 1/k[OH] = 1 \text{ day.}$$

After evaluating the available data (Morris et al. 1971, Morris and Niki 1971, Niki et al. 1978, Atkinson and Pitts 1978, Kerr and Sheppard 1981, Semmes et al. 1985, and Michael et al. 1985) concerning the kinetics and mechanism of the reaction of the hydroxyl radical with acetaldehyde, Atkinson (1989) recommended a rate constant of:

$$k = 1.58 \times 10^{-11} cm^3 \text{ molecule}^{-1} \text{second}^{-1} \pm 20\% \text{ at } 298^{o}K.$$

Prinn et al. (1987) estimated the concentration of hydroxyl radicals in the atmosphere to be 1.5×10^6 molecules cm^{-3} x as a 12 hour daytime average.

While definite product and mechanistic data are not available, peroxyacetyl nitrate (PAN) has been observed as a product of the reaction of the hydroxyl radical with acetaldehyde in room temperature air when nitrogen oxides are present (Atkinson and Lloyd 1984). This shows that, at room temperature, this reaction must proceed via overall hydrogen atom abstraction from the -CHO group (hydrogen atom abstraction from the $-CH_3$ group is expected to be of minimal importance at room temperature) (Atkinson 1987). The initial hydrogen atom abstraction forms the acetyl radical which, exclusively under tropospheric conditions, combines with oxygen to yield the acetyl peroxy radical. The acetyl peroxy radical then reacts with nitrogen oxides to create PAN (Atkinson 1990) (see Section C, Fate of Atmospheric Acetaldehyde for the chemical reaction).

b. Photolysis

Photolysis of acetaldehyde is the chemical decomposition of atmospheric acetaldehyde induced by sunlight. The rate of photolysis depends on three variables which are, in turn, wave-length dependent:

absorption cross-section (σ)[1], photolysis quantum yield (ϕ)[2], and radiation intensity (J)[3].

$$1/\tau= k= \begin{matrix}\text{~800 nm}\\ \text{~290 nm}\end{matrix} \alpha_{\lambda}\phi_{\lambda}J_{\lambda}d\lambda= 6.6 \text{ days at zenith angle } 0.^{\circ} \quad 4$$

Photolysis is of secondary importance as an atmospheric acetaldehyde removal mechanism (τ= 6.6 days) when compared to the acetaldehyde-hydroxyl radical reaction (τ= 1 day) (Horowitz and Calvert 1982, Meyrahn et al. 1982). The photodissociation is expected to proceed in the following manner:

$$CH_3CHO + h\nu \rightarrow CH_4 + CO$$

and

$$CH_3CHO + h\nu \rightarrow CH_3 + HCO$$

c. Reaction with Ozone (O_3), Nitrate (NO_3) Radicals, and Hydroperoxyl (HO_2) Radicals

The atmospheric lifetime of acetaldehyde resulting from oxidation by ozone, nitrate radicals, and hydroperoxyl radicals is estimated as >4.5 years, 35 days, and 3 years respectively. The ozone and hydroperoxyl radical reactions proceed during both day and night hours while the nitrate radical reaction occurs at night. These lifetimes were determined by the rate constants and concentration of reactants associated with each reaction (τ= k[B] where B is the atmospheric oxidant) (Atkinson 1989).

2. <u>Physical Removal from the Atmosphere</u>

Atkinson evaluated the limited available data on physical removal processes and found that dry and wet deposition are expected to only slightly increase the atmospheric removal of acetaldehyde above that of the acetaldehyde-hydroxyl reaction (Atkinson 1988 and 1989, Bidleman 1988).

Dry deposition removes organic gases by absorption or adsorption to snow or moist surfaces. Because dry deposition occurs only at ground surfaces, it plays a limited role in removing acetaldehyde from the atmosphere.

Since acetaldehyde dissolves in aqueous solutions, it is reasonable to assume that the compound is subject to wet deposition by incorporation into clouds, rain, and fog. However, the degree of acetaldehyde hydration calculated using the Henry's law coefficient ($H^* = 11.4 \pm 0.4$ mol atm^{-1} at 298°K) of Betterton and Hoffmann (1988) is fairly small (1.4) when compared to that of formaldehyde (2.3×10^3). In addition, gas-phase organic compounds which are efficiently rained out have a washout ratio (concentration in rain/concentration in air) of 10^5 to 10^6. Atkinson (1989) calculated an acetaldehyde washout ratio of 28 at 298°K using the H* value of Betterton and Hoffmann (1988). This washout ratio compares favorably with the washout ratio of 37 reported earlier by Buttery et al. (1969). The episodic nature of precipitation events and the estimated low washout ratio of acetaldehyde indicate that wet deposition is not a significant removal mechanism (Atkinson 1989).

C. FATE

As mentioned in Section B, Persistence of Acetaldehyde, reaction with the hydroxyl radical and, to a lesser extent, photolysis are the major tropospheric chemical loss processes. The reaction with the hydroxyl radical and photolysis occur only during daytime hours with the acetaldehyde-hydroxyl radical reaction serving as the driving force for the estimated overall lifetime of acetaldehyde. Wet deposition may lead to a shorter tropospheric lifetime, but wet deposition is episodic in nature and expected to be of minor importance due to the relatively low acetaldehyde washout ratio. The one-day lifetime calculated from the acetaldehyde-hydroxyl radical reaction can be regarded as the tropospheric lifetime of acetaldehyde with somewhat faster removal expected during rain or fog (Grosjean and Wright 1983).

The reactions subsequent to the hydroxyl radical reaction begin with hydrogen atom abstraction resulting in the formation of the acetyl radical (CH_3CO)

$$OH + CH_3CHO \rightarrow H_2O + CH_3CO.$$

Exclusively under tropospheric conditions, the acetyl radical is rapidly oxidized to the acetyl peroxy radical ($CH_3C(O)OO$)

$$CH_3CO + O_2 \rightarrow CH_3C(O)OO.$$

The acetyl peroxy radical reacts with nitric oxide and nitrogen dioxide, with the latter reaction forming peroxy acetyl nitrate ($CH_3C(O)OONO_2$) (PAN)

$$\begin{tikzcd} CH_3C(O)OO + NO \arrow[r] & CH_3C(O)O + NO_2 \arrow[d] \\ & CH_3 + CO_2 \end{tikzcd}$$

$$CH_3C(O)OO + NO_2 \rightleftarrows \underset{\text{(PAN)}}{CH_3C(O)OONO_2} .$$

Additionally, in a "clean" troposphere with low levels of nitrogen oxides, the reactions of the acetyl peroxy radical with hydroperoxyl and methyl peroxy (CH_3O_2) radicals may result in end products of formaldehyde (HCHO) and acetic acid (CH_3COOH) (Atkinson 1989, 1990).

References for Chapter V

Atkinson, R. and J.N. Pitts, Jr. (1978) "Kinetics of the Reactions of the OH radical with HCHO and CH_3CHO over the Temperature Range 299-426°K", J. Chem. Phys., Vol. 68, pp. 3581-3584.

Atkinson, R. and A. C. Lloyd (1984) "Evaluation of Kinetics and Mechanistic Data for Modeling of Photochemical Smog", J. Phys. Chem. Ref. Data, Vol. 13, pp. 315-444.

Atkinson, R. and W. Carter (1984) "Kinetics and Mechanisms of the Gas-phase Reactions of Ozone with Organic Compounds under Atmospheric Conditions", Chemical Reviews, Vol. 84, pp. 437-470.

Atkinson, R. (1987) "A Structure-activity Relationship for the Estimation of Rate Constants for the Gas-phase Reactions of OH Radicals with Organic Compounds", International Journal of Chemical Kinetics, Vol. 19, pp. 799-828.

Atkinson, R. (1988) "Atmospheric Transformations of Automotive Emissions", Air Pollution, the Automobile, and Public Health, A.Y. Watson, R.R. Bates and D. Kennedy (editors), National Academy Press, Washington, D.C., pp. 99-132.

Atkinson, R., (1989) "Atmospheric Lifetime and Fate of Acetaldehyde", California Air Resources Board Contract A732-107, Statewide Air Pollution Research Center, University of California, Riverside, California.

Atkinson, R. (1990) "Gas-phase Tropospheric Chemistry of Organic Compounds: a Review", Atmospheric Environment, Vol. 24 A, pp. 1-41.

Atkinson, R., (1989b) "Atmospheric Lifetime and Fate of Acetaldehyde", California Air Resources Board Contract A732-107, Statewide Air Pollution Research Center, University of California, Riverside, California.

Betterton, E.A. and M. R. Hoffmann (1988) "Henry's Law Constants of Some Environmentally Important Aldehydes", Environmental Science and Technology, Vol. 22, pp. 361-367.

Bidleman, T.F. (1988) "Atmospheric Processes", Environmental Science and Technology, Vol. 22, pp. 361-367.

Buttery, R.G., Ling, L.C. and Guadagni, D.G. (1969) "Volatilities of Aldehydes, Ketones, and Esters in Dilute Water Solution", Journal of Agriculture and Food Chemistry, Vol. 17, pp. 385-389.

Dlugokencky, E. and C. J. Howard (1989) "Studies of NO_3 Radical Reactions with Some Atmospheric Organic Compounds at Low Pressures", Journal of Physical Chemistry, Vol. 93, pp. 1091-1096.

Finlayson-Pitts, B. J. and J. N. Pitts, Jr. (1986) "Atmospheric Chemistry: Fundamentals and Experimental Techniques", John Wiley and Sons, Inc., New York, NY, p. 8.

Grosjean, D., Swanson, R.D. and Ellis, C. (1983) "Carbonyls in Los Angeles Air: Contribution of Direct Emissions and Photochemistry", The Science of the Total Environment, Vol. 29, pp. 65-85.

Grosjean, D. and B. Wright (1983) "Carbonyls in Urban Fog, Ice Fog, Cloudwater, and Rainwater", Atmospheric Environment, Vol. 17, pp. 2093-2096.

Hard, T. M., O'Brien, R. J., Chan, C. Y. and Mehrabzaden, A. A. (1984) "Tropospheric Free Radical Determination by FAGE", Environmental Science and Technology, Vol. 18, pp. 768-777.

Horowitz, A. and Calvert, J. G. (1982) "Wavelength Dependence of the Primary Processes in Acetaldehyde Photolysis", Journal of Physical Chemistry, Vol. 86, pp. 3105-3114.

Kerr, J. A. and D. W. Sheppard (1981) "Kinetics of the Reactions of Hydroxyl Radicals with Aldehydes Studied under Atmospheric Conditions", Environmental Science and Technology, Vol. 15, pp. 960-963.

Logan, J. A. (1985) "Tropospheric Ozone: Seasonal Behavior, Trends, and Anthropogenic Influence", Journal of Geophysical Research, Vol. 90, pp. 10463-10482.

Meyrahn, H., Moortgat, G. K. and Warneck, P. K. (1982) "The Photolysis of Acetaldehyde under Atmospheric Conditions", 15th Informal Conference on Photochemistry, Stanford University, Stanford, CA, June 27-July 1, 1982.

Michael, J. V., D. G. Kell and Klemm, R. B. (1985), "Rate Constants for the Reaction of Hydroxyl Radicals with Acetaldehyde from 244-528°K", J. Chem. Phys.", Vol. 83, pp. 1630-1636.

Moortgat, G. K., Burrows, J. P., Schneider, W., Tyndall, G. S. and Cox, R. A. (1987), "A Study of the HO_2 + CH_3CHO Reaction in the Photolysis of CH_3CHO, and its Consequences for Atmospheric Chemistry", Proc. 4th European Symposium on the Physico-chemical Behavior of Atmospheric Pollutants, Reidel Publishing Company, Dordrecht, Holland, pp. 271-281.

Morris, E. D., Jr. and H. Niki (1971) "Reactivity of Hydroxyl Radicals with Olefins", Journal of Physical Chemistry, Vol. 75, pp. 3640-3641.

Morris, E. D., Jr., Stedman, D. H. and Niki, H. (1971) "Mass Spectometric Study of the Reactions of the Hydroxyl Radical with Ethylene, Propylene, and Acetaldehyde in a Discharge-flow System", Journal of American Chemical Society, Vol. 93, pp. 3570-3572.

Niki, H., Maker, P. D., Savage, C. M. and Breitenbach, L. P. (1978) "Relative Rate Constants for the Reaction of Hydroxyl Radical with Aldehydes", Journal of Physical Chemistry, Vol. 82, pp. 132-134.

Platt, U. F., Winer, A. M., Biermann, H. W., Atkinson, R. and Pitts, J. N., Jr., (1984) "Measurement of Nitrate Radical Concentrations in Continental Air", Environmental Science and Technology, Vol. 18, pp. 365-369.

Prinn, R., Cunnold, D., Rasmussen, R., Simmonds, P., Alyea, F., Crawford, A., Fraser, P. and Rosen, R., (1987) "Atmospheric Trends in Methylchloroform and the Global Average for the Hydroxyl Radical", Science, Vol. 238, pp. 945-950.

Semmes, D. H., Ravishankara, A. R., Gump-Perkins, C. A. and Wine, P. H., (1985) "Kinetics of the Reactions of Hydroxyl Radical with Aliphatic Aldehydes", International Journal of Chemical Kinetics, Vol. 17, pp. 303-313.

APPENDIX A

PHOTOCHEMICALLY FORMED ACETALDEHYDE AS A PERCENT OF AMBIENT CONCENTRATION

State of California

M E M O R A N D U M

To : Beth Schwehr, Manager
Toxics Emission Inventory Section
Technical Support Division

Date : May 31, 1991

Subject : Estimates of Secondary Ambient Acetaldehyde

Dr. Kit Wagner, Manager
Control Strategy Modeling Section
Technical Support Division

From : **Air Resources Board**

This memo is to clarify certain portions of our estimates of secondary acetaldehyde in the South Coast Air Basin. All estimates of the secondary component of ambient acetaldehyde concentrations are for the South Coast Air Basin. These estimates are based on simulations for August 28, 1987 using the Urban Airshed Model with the Carbon Bond 4 chemistry with and without primary emissions of acetaldehyde. This day represents meteorological conditions conducive to high ozone concentrations over the South Coast Air Basin and thus probably represents an upper limit for secondary concentrations of acetaldehyde.

We selected 32 monitoring sites throughout the basin with the intent of focusing on areas of high population. The average daily concentrations of acetaldehyde were estimated for secondary, primary, and boundary/initial condition components. The boundary/initial condition component of the acetaldehyde concentrations are very small, averaging about 2 percent or less of the total concentration. The secondary fractions of the ground level acetaldehyde concentrations are shown in Table 1.

Table 1

Secondary Acetaldehyde
Percent of Ambient Concentration
Averaged Over 32 Sites

Daily	Daytime	Nighttime
55.6	67.0	41.1

These statistics were computed from the 32 sites shown in Attachment 1 to this memorandum. The daily ratio is based on the overall 24-hour acetaldehyde concentrations of secondary to total acetaldehyde concentrations. Daytime includes hours 6am through 7pm PDT (13 hours), while nighttime is for 7pm through 6am PDT (11 hours).

The daily average secondary acetaldehyde is 56% over the 32 sites we have considered. Because we used a summer high ozone day this is likely to significantly overpredict the annual average secondary acetaldehyde.

Table 2 shows the acetaldehyde concentrations averaged over the 32 sites for each case in Table 1.

Table 2

Acetaldehyde Concentrations (PPB) for August 28, 1987
(Averaged Over 32 Sites)

	Daylight	Nighttime	Daily
Secondary	5.31	2.98	4.24
Total	7.93	7.25	7.62

The Carbon Bond 4 mechanism represents the acetaldehyde mechanism and rate constants with the model species ALD2. Acetaldehyde and certain other higher carbonyls are grouped together as ALD2. Our results actually represent ALD2. These results should be considered approximate in terms of acetaldehyde. We also wish to point out that this model underpredicts ozone and hydrocarbons.

cc: Terry McGuire
Don McNerny
Linda Murchison
Joan Denton (SSD)
Chris Nguyen

PDA//B48092

ATTACHMENT 1

Site Specific Modeled Acetaldehyde Concentrations
August 28, 1987
South Coast Air Basin
(ppb)

SITE	PRIMARY	SECONDARY
Anaheim	3.2	3.0
Azusa	4.0	5.3
Beverly Hills	4.4	4.2
Burbank	4.1	6.1
Downtown LA	5.1	5.0
Claremont	3.4	6.0
Costa Mesa	1.8	1.7
El Monte	4.3	4.7
Fontana	3.2	7.5
Glendora	3.5	6.1
Hawthorne	2.0	1.5
La Habra	4.3	5.0
Lancaster	0.4	3.1
Long Beach	1.2	1.7
Los Alamitos	2.7	2.4
Lynnwood	2.7	2.6
Newhall	3.8	7.3
Norco	3.9	5.3
Pasadena	5.0	6.0
Pico Rivera	4.0	3.9
Palm Springs	1.0	3.4
Pomoma	3.6	5.8
Redlands	3.9	10.6
Reseda	4.6	7.4
Riverside	3.2	6.9
San Bernardino	4.3	10.3
Thousand Oaks	2.4	3.3
Torrance	9.1	2.5
Upland	2.7	6.1
Victorville	0.6	4.9
Whittier	4.0	4.3
West LA	2.1	3.0

2. CALCULATIONS TO ESTIMATE PHOTOCHEMICALLY FORMED ACETALDEHYDE

This section addresses the calculation the staff used to estimate the tons of acetaldehyde per year formed by photo-chemical oxidation.

The term "direct" or "primary" acetaldehyde refers to acetaldehyde that is a combustion product or directly emitted from a source. Examples of these emissions include acetaldehyde from tailpipes, smokestacks and burning fields. Table III-1 shows California direct and secondary acetaldehyde emissions in 1987. Direct emissions from mobile and stationary sources ranged from 11,444 to 23,902 tons per year.

Acetaldehyde can also be formed in the atmosphere by photochemical oxidation of certain hydrocarbons from these sources and is called "secondary" acetaldehyde. An air pollution model (Appendix A.1). predicts that an average 55.6 percent of acetaldehyde is secondarily formed during a 24 hour period on a hot summer day. Thus, if the photochemically formed acetaldehyde is 55.6 percent then the direct acetaldehyde would constitute 44.4 percent of the total acetaldehyde. The photo-oxidation acetaldehyde product can be calculated if formed in the same geographical location and time period.

To estimate the range of secondary acetaldehyde emissions the following calculations were used. The tonnage is not to be taken as concise numbers but rather is provided to put a perspective on the contribution of photochemical-oxidation to total acetaldehyde in ambient air.

EMS: photochemical acetaldehyde product, tons/year
44.4: percent of primary emissions of acetaldehyde
55.6: percent of secondary emissions of acetaldehyde
11,444: acetaldehyde low range direct emissions, tons/year
23,902: acetaldehyde high range emissions, tons/year

Lower end of range: $\frac{44.4}{11,444} = \frac{55.6}{EMS}$;

EMS = 14,330 secondary acetaldehyde, tons/year
(approximately 14,000 tons/year)

Upper end of range: $\frac{44.4}{23,902} = \frac{55.6}{EMS}$;

EMS = 29,931 secondary acetaldehyde, tons/year
(approximately 30,000 tons/year)

APPENDIX B

EMISSION ESTIMATES

EMISSION ESTIMATES

I. Mobile Sources

1. On-Road Motor Vehicles

In general, the lower limit of acetaldehyde emissions from on-road motor vehicles is estimated as follows:

$$EMS = TOG * (THC/TOG) * F \qquad (1)$$

Where:

EMS : acetaldehyde emissions, tons/year.
TOG : total organic gas emissions for specific vehicular classes, tons/year.
THC : total hydrocarbon emissions, tons/year.
F : fraction of acetaldehyde to total hydrocarbons, dimensionless.

Acetaldehyde emissions from all on-road motor vehicles along with the appropriate fractions of acetaldehyde to total hydrocarbons (THC) are tabulated in Table A-1. This Table lists the lower limits of acetaldehyde emissions for different classes of vehicles.

The methodology to estimate the upper limits of acetaldehyde emission from on-road motor vehicles is explained in the main text. Table A-2 lists the upper limits of acetaldehyde emissions along with the TOG emissions and the fractions of acetaldehyde to TOG from on-road vehicular exhaust.

2. Other Mobile Sources

The ARB has compiled volatile organic compound (VOC) profiles for a number of fuel oil combustion processes (ARB, 1989). Based on these profiles and the total organic gas (TOG) emissions from off-road vehicles, trains, ships, aircraft, utility and mobile equipment, estimates of acetaldehyde emissions from these sources are as follows:

$$EMS = TOG * FA \qquad (2)$$

Where:

EMS : acetaldehyde emissions, tons/year.
TOG : total organic gas emissions for specific category, tons/year.
FA : fraction of acetaldehyde to TOG, dimensionless.

Table A-3 lists acetaldehyde emissions along with TOG emissions for other mobile sources in California.

Table A-1

Lower Limits of Acetaldehyde Emissions from On-Road Motor Vehicles for 1987

	THC[1]	Acetaldehyde[1]	
	(tons/yr)	(% of THC)	(tons/yr)
Light-Duty Passenger Cars			
Non Catalyst	113,281	0.44	498
Catalyst	125,324	0.31	389
Diesel	1,560	0.54	8
Light- and Medium-Duty Trucks			
Non Catalyst	40,122	0.44	177
Catalyst	53,365	0.31	165
Diesel	405	1.11	4
Heavy-Duty Trucks			
Non Catalyst	17,411	0.54	94
Catalyst	161	0.38	1
Diesel	25,143	1.11	279
Urban Buses	1,067	1.11	12
Motorcycles	5,060	0.44	22
Total			1,649

---

[1] ARB, 1990a. Emissions have been rounded to the nearest whole number.

Table A-2
Upper Limits of Acetaldehyde Emissions from On-Road Motor Vehicles for 1987

	TOG[1] (tons/yr)	Acetaldehyde[2] Fraction	Acetaldehyde[1] (tons/yr)
Light-Duty Passenger Cars			
Non Catalyst	119,963	0.0190	2,280
Catalyst	139,046	0.0059	820
Diesel	1,810	0.0296	54
Light- and Medium-Duty Trucks			
Non Catalyst	42,488	0.0190	807
Catalyst	59,320	0.0059	350
Diesel	469	0.0296	14
Heavy-Duty Trucks			
Non Catalyst	20,250	0.0190	385
Catalyst	178	0.0059	1
Diesel	29,244	0.0296	865
Urban Buses	1,243	0.0296	37
Motorcycles	5,359	0.0190	102
Total			5,715

Table A-3
Acetaldehyde Emissions from Other Mobile Sources for 1987

	TOG[1] (tons/year)	Acetaldehyde[2] Fraction	Acetaldehyde[1] (tons/year)
Off-Road Vehicles			
Diesel Combustion	1,652	0.0296	49
Gasoline Combustion	38,407	0.0190	730
Trains	10,819	0.0296	320
Ships	2,533	0.0296	75[3]
Aircraft	1,622	0.0190	31
Mobile and Utility Equipment			
Diesel Combustion	9,971	0.0296	295
Gasoline & LPG Combustion	30,494	0.0190	579
Total			2,079

---

[1] ARB, 1990b; and ARB, 1990c. Emissions have been rounded to the nearest whole number.

[2] ARB, 1989.

[3] Approximately 21 % of this estimate is for activities occurring in the Outer Continental Shelf area.

III. Stationary Area Sources

For stationary area sources, the staff has used a methodology similar to that used in estimating acetaldehyde emissions from fuel combustion sources. The staff has used the amount of material burned and the appropriate emission factors to estimate acetaldehyde emissions from stationary area sources. These estimates are reported in Table A-6 as follows:

Table A-6
Acetaldehyde Emissions from Stationary Area Sources

Material Burned	Process Rate[2]	Emission Factor	Emissions[1] (tons/year)
Wood	1.12×10^6 tons	0.20 lb/ton	110 - 660[3]
Diesel	6.29×10^6 gallons	Not applicable	43[4]
Oil[5]	1.21×10^6 tons	0.0016 - 0.0032 lb/ton	0.97 - 1.9
Natural Gas	8.38×10^{11} ft^3	$(1.7 - 3.4) \times 10^{-8}$ lb/ft^3	7 - 14
Ag. Burning	2.81×10^6tons	0.73 - 1.5 lb/ton	1,030 - 2,110
Manag't Burning	3.32×10^6 tons	0.73 - 1.5 lb/ton	1,210 - 2,490
Open Burning	2.38×10^5 tons	0.73 - 1.5 lb/ton	87 - 178
Wildfires	1.23×10^7 tons	0.73 - 1.5 lb/ton	4,490 - 9,220
Total			6,980 - 14,700

---

1. These numbers may not exactly equal the process rates time the emission factors due to rounding in the original calculations (at most three significant figures).

2. ARB (1990f) These process rates have been rounded off to at most three significant figures.

3. The upper number was estimated using TOG emissions and VOC speciation profile for residential wood combustion.

4. Calculated from TOG emissions and the fraction of acetaldehyde in TOG. A small portion (less than 1%) of this emission estimate represents emissions for activities occurring in the Outer Continental Shelf (OCS) area.

5. This category also includes 25% from unspecified liquid material combustion sources.

II. Stationary Point Sources

1. Fuel Combustion

In general, acetaldehyde emissions from all fuel combustion sources were estimated as follows:

$$EMS = PR * EF \qquad (3)$$

Where:

EMS = acetaldehyde emissions, tons/year.

PR = amount of coal, oil, or gases burned; tons, gallons, or million cubic feet per year.

EF = emission factor; lbs per ton, gallon, or million cubic feet.

Acetaldehyde emissions for each of the fuel combustion types, along with the process rates and emission factors are tabulated in Table A-4 as follows:

Table A-4
Acetaldehyde Emissions from Fuel Combustion Sources

Fuel Type	Process Rate[1]	Emission Factor[2]	Emissions (tons/year)
Coal	1.6×10^6 tons	0.0014 - 0.0029 lb/ton	1 - 2
Diesel	3.24×10^6 gallons	Not Applicable	1[3]
Gasoline	4.53×10^6 gallons	Not Applicable	1000
Natural Gas	1.8×10^{12} ft^3	$(1.7 - 3.4) \times 10^{-8}$ lb/ft^3	15 - 31
Oil	1.26×10^7 tons	0.0016 - 0.0032 lb/ton	10 - 20
Wood	5.63×10^6 tons	0.20 lb/ton	563

---

[1] ARB, 1990c.
[2] MRI, 1987.
[3] ARB, 1990h. Calculated from TOG emissions and the fraction of acetaldehyde in TOG.

2. Refineries - Processing

Based on the "Annual Refining Survey" of the Oil and Gas Journal, California refineries had a rated charge capacity of approximately 2.28 million barrels per calendar day of crude oil in 1988 (OGJ, 1989). Catalytic cracking units had a rated charge capacity of 645,500 bbls per stream day or 223,827 thousand bbls per calendar year and thermal operations (assuming coking operations) had a rated charge capacity of 503,200 bbls per stream day or 174,485 thousand bbls per calendar year in 1988 (OGJ, 1989). Estimates of acetaldehyde emissions for refineries are as follows:

$$EMS = RC * EF \qquad (4)$$

Where:

EMS = acetaldehyde emissions, tons/year.
RC = rated charge capacity, bbls/yr.
EF = acetaldehyde emission factor, lbs/1000 bbls.

Acetaldehyde emission factors are different depending on the refining processes; thus, acetaldehyde emissions are different from these processes. Table A-5 lists acetaldehyde emissions from each refining process, along with the charge capacities and the acetaldehyde emission factors.

Considering the possibility of additional acetaldehyde emissions from other operations such as catalytic reforming, catalytic hydrorefining and catalytic hydrotreating, a range of estimates includes acetaldehyde emissions from petroleum refining processes. To account for these other operations, the process rates for all catalytic units as reported in the OGJ have been added to estimate maximum process rates for catalytic cracking units. This process rate is then used to estimate the high end of the range shown in Table A-5 for acetaldehyde emissions from catalytic cracking units in refineries.

Table A-5

Acetaldehyde Emissions from Refinery Processing in California

Source	Charge capacity[1] (1000 bbls/yr)		EF (lbs/1000 bbls)	Acetaldehyde Emissions (tons/yr)	
	Low	High		Low	High
Catalytic Cracking	223,827	1,062,945[2]	1.40	157	744
Thermal Operations[3]	174,485	174,485	0.34	30	30
Total[4]				190	770

---

1. OGJ, 1989. Charge capacities are expressed as the volume of crude oil that the facility can process per calendar year. These numbers were estimated based on the rated charge capacities per stream day. The OGJ assumed that the capacity per calendar day equals approximately 95% of the capacity per stream day.

2. This number was estimated by adding all process rates for catalytic units as reported in the OGJ. Specifically, 223,827 thousand bbls of oil were processed in catalytic cracking units; 194,700 thousand bbls in catalytic reforming units; 133,499 thousand bbls in catalytic hydrorefining units, and 328,875 thousand bbls in catalytic hydrotreating units.

3. Assumed to be coking operations.

4. The totals have been rounded to at most two significant figures.

3. Coffee Roasting

In 1987, coffee makers in California roasted approximately 279,000 tons of green coffee beans. Most of these facilities use an after burner as control equipment (ARB, 1990). Acetaldehyde emissions from coffee bean roasting in California are estimated as follows:

$$EMS = PR * EF \qquad (5)$$

Where:

EMS = acetaldehyde emissions, tons per year.
PR = amount of green beans roasted, tons/year.
EF = controlled emission factor, ton acetaldehyde per ton green beans.

$$EMS = 278{,}970 \text{ tons/yr} * 8.2 \times 10^{-6} \text{ ton/ton}$$

$$= 2.3 \text{ tons acetaldehyde per year.}$$

APPENDIX C

ANALYSIS OF ACETALDEHYDE

The following procedure is used to analyze acetaldehyde as well as formaldehyde.

State of California
Air Resources Board
Monitoring & Laboratory Division

Standard Operating Procedure for the Determination of Formaldehyde in Ambient Air

1.0 Scope

This is a high performance liquid chromatographic (HPLC) method for the determination of formaldehyde in the ambient air utilizing solid adsorbent. The method was developed from EPA Method TO11.

2.0 Summary of Method

2.1 Ambient air is drawn through Sep-Pak chromatographic grade ~~silica gel~~ cartridges. The cartridges are coated with acidified 2,4-dinitrophenylhydrazine (DNPH). Sampling rate is dependent upon carbonyl concentration but is estimated to be 0.7L per minute maximum for a 24-hour period. During sampling, formaldehyde reacts with the DNPH to form a DNPH-formaldehyde derivative (DNPH-HCHO).

2.2 The DNPH-HCHO is eluted from the sampling cartridges using acetonitrile (ACN) and is quanitified using reverse-phase HPLC with ultraviolet absorption detection at 360 nm.

3.0 Interferences/Limitations

3.1 Since this procedure is written for the sampling and analysis of formaldehyde, interferences resulting from certain isomeric aldehydes that may be unresolved by the HPLC system when analyzing for other aldehydes is not a factor.

3.2 Formaldehyde contamination of the DNPH reagent is a frequently encountered problem. The DNPH must be purified by multiple recrystallations in UV grade ACN. Recrystallization is accomplished at 40–60°C by slow evaporation of the solvent to maximize crystal size. The purified DNPH crystals are stored under UV grade acetonitrile until use. Impurity levels in the DNPH are determined by HPLC prior to use.

4.0 Apparatus

4.1 A gradient HPLC system consisting of a mobile phase reservoir; high pressure pumps; an injection valve or automatic sampler; a C-18 reverse phase column 3.9 mmx 30 cm; a variable wavelength UV detector operating at 360 nM; and a data system.

4.2 Sampling system - a XonTech Module 920 Multi-media sampler outfitted with a sampling head configured to hold the adsorbant Sep-Pak cartridges and capable of sampling at a flow rate of between 0.5 and 2.0 LPM.

4.3 Supelco 4 mL glass desorption vials with teflon lined screw caps. Sun brokers 1 mL autosampler "Sun Vial" with polyethylene cap septum.

4.4 Sample shaker/desorber and sample racks.

4.5 Gelman AcroPrep syringeless 0.45 nm PTFE membrane sample filter units.

4.6 Filtration and degassing system for mobile phase solvents such as Waters Part #85124.

4.7 Various volumetric pipets and flasks and graduated cylinders.

4.8 Special glass apparatus for rinsing, storing and dispensing saturated DNPH stock reagent (see Figure #<u>1</u>).

4.9 Hotplate.

4.10 Polyethylene gloves - used to handle the treated cartridges.

5.0 <u>Reagents</u>

5.1 2,4-Dinitrophenylhydrazine (DNPH) - Aldrich Chemical (Catalog #D19,930-3) or equivalent. Recrystallize at least twice with UV grade acetonitrile before use.

5.2 Acetonitrile and water mobile phase solvents, HPLC grade such as Burdick & Jackson Product #016 and Baker #4218-3, respectively.

5.3 Hydrochloric acid - analytical grade.

5.4 Formaldehyde - analytical grade.

5.5 Methanol or Ethanol - analytical grade.

5.6 Sep-Pak silica gel cartridges - Waters Associates, 34 Maple Street, Milford, MA - Product #51900

5.7 Calibration Standards are prepared in ACN from the DNPH-formaldehyde derivative. A stock solution of 100 mg/L is prepared by dissolving 10 mg of solid derivative in 100 mL of ACN. Calibration standards spanning the range of interest are prepared from the stock solution.

6.0 Preparation of Reagents & Cartridges

6.1 Purification of 2,4-Dinitrophenylhydrazine (DNPH) [Note: This procedure should be performed under a properly ventilated hood.]

6.1.1 Prepare a supersaturated solution of DNPH by boiling excess DNPH in 200 mL of acetonitrile for approximately one hour.

6.1.2 After one hour, remove and transfer the supernatant to a covered beaker on a hot plate and allow gradual cooling to 40–60°C.

6.1.3 Let covered beaker heat at this temperature range, allowing 95% of solvent to evaporate slowly.

6.1.4 Decant solution to waste, and rinse crystals twice with three times their apparent volume of acetonitrile. [CAUTION: Various health effects are resultant from the inhalation of acetonitrile. At 500 ppm in air, brief inhalation has produced nose and throat irritation. At 160 ppm, inhalation for 4 hours has caused flushing of the face (2 hour delay after exposure) and bronchial tightness (5 hour delay). Heavier exposures have produced systemic effects with symptoms ranging from headache, nausea, and lassitude to vomiting, chest or abdominal pain, respiratory depression, extreme weakness, stupor, convulsions and death (dependent upon concentration and time)].

6.1.5 Transfer crystals to another clean beaker, add 200 mL of acetonitrile, heat to boiling, and again let crystals grow slowly at 40–60°C until 95% of the solvent has evaporated.

6.1.6 Repeat rinsing process as described in Section 6.1.4.

6.1.7 Take an aliquot of the second rinse, dilute 10 times with acetonitrile, acidify with 1 mL of 3.8 M perchloric acid per 100 mL of DNPH solution, and analyze by HPLC. An ideal impurity level is shown in Figure 2.

6.1.8 The impurity level of DNPH should be below the sensitivity (ppb, v/v) level indicated in Table 1 for the anticipated sample volume. If the impurity level is not acceptable for intended sampling application, repeat recrystallization.

6.1.9 Transfer the purified crystals to an all-glass reagent bottle, add 200 mL of ACN, stopper, shake gently, and let stand overnight. Analyze supernatant by HPLC. The purity level should be comparable to that shown in Figure #__2__.

6.2 Preparation of DNPH-Formaldehyde Derivative

6.2.1 Titrate a saturated solution of DNPH in 2N HC1 with formaldehyde (other aldehydes may be used if their detection is desirable, see Fig.3 for a typical chromatagram of the various DNPH-Carbonyl derivatives.).

6.2.2 Filter the colored precipitate, wash with 2N HC1 and water and let precipitate air dry.

6.2.3 Check the purity of the DNPH-formaldehyde derivative by melting point determination table or HPLC analysis.

6.3 Preparation of DNPH-formaldehyde Standards

6.3.1 Prepare a standard stock solution of the DNPH-formaldehyde derivative by dissolving accurately weighed amounts in acetonitrile.

6.3.2 Prepare a working calibration standard mix from the standard stock solution. The concentration of the DNPH-formaldehyde compound in the standard mix solutions should be adjusted to reflect relative distribution in a real sample. [Note: Individual stock solutions of approximately 100 mg/L are prepared by dissolving 10 mg of the solid derivative in 100 mL of acetonitrile. The individual solution is used to prepare calibration standards containing the derivative of interest at concentrations of 0.1-10 mg/L, which spans the concentration of interest for most ambient air work.]

6.3.3 Store all standard solutions in a refrigerator. They are stable for several months.

6.4 Preparation of DNPH-Coated Cartridges

[Note: This procedure must be performed in an atmosphere with a very low aldehyde background. All glassware and plastic ware must be scrupulously cleaned and rinsed with deionized water and aldehyde free acetonitrile. Contact of reagents with laboratory air must be minimized. Polyethylene gloves must be worn when handling the cartridges.]

6.4.1 DNPH Coating Solution

6.4.1.1 Dilute 25 mL of saturated DNPH stock solution to 100 mL with acetonitrile in a reagent bottle equipped with a positive displacement dispenser.

6.4.1.2 Acidify with 1.0 mL of concentrated HC1. [Note: The atmosphere above the acidified solution should preferably be filtered through a DNPH-coated silica gel cartridge to minimize contamination from laboratory air.]

6.4.1.3 Prime the dispenser and slowly dispense 10-20 mL to waste.

6.4.1.4 The impurity level should be similar to that shown in Figure 2.

6.4.2 Coating of Sep-Pak Cartridges

6.4.2.1 Open the Sep-PAk package, connect the short end to a 10-mL syringe, and place it in the syringe rack. [Note: Prepare as many cartridges and syringes as possible.]

6.4.2.2 Using a positive displacement repetitive pipet, add 10 mL of acetonitrile to each of the syringes.

6.4.2.3 Let liquid drain to waste by gravity. [Note: Remove any air bubbles that may be trapped between the syringe and the silica cartridge by displacing them with the acetonitrile in the syringe.]

6.4.2.4 Set the repetitive dispenser containing the acidified DNPH coating solution to dispense 10mL into the cartridges.

6.4.2.5 Once the effluent flow at the outlet of the cartridge has stopped, dispense 10mL of the DNPH coating reagent into each of the syringes.

6.4.2.6 Let the coating reagent drain by gravity through the cartridge until flow at the other end of the cartridge stops.

6.4.2.7 Wipe the excess liquid at the outlet of each of the cartridges with clean tissue paper.

6.4.2.8 Remove the cartridges from the syringes and connect the short ends to the Luer ports of the drying manifold.

6.4.2.9 Connect a DNPH-coated Sep-Pak cartridge to the outlet port of the drying manifold. This "guard cartridge" will serve to remove any trace of formaldehyde in the nitrogen gas supply.

6.4.2.10 Pass nitrogen through each of the cartridges at about 300-400 mL/min for 5-10 minutes.

6.4.2.11 After drying, arrange the Sep-Paks in pairs using a one inch piece of 3/16" ID Tygon tubing. Make sure the Sep-Pak cartridge ends butt together. The first cartridge is the primary (sampling cartridge) while the second is used to detect sample breakthrough. Cap the two exposed cartridge ends with 3/16" ID column caps.

6.4.2.12 Store cartridges in a screw cap glass culture tube in a refrigerator until use.

6.4.2.13 Before transport, remove the screw-capped glass culture tubes containing the adsorbent tube pairs from the refrigerator and place the tubes in a friction-top metal tube and the metal tube in fiberboard screw cap mailing case.

7.0 Sampling

7.1 A XonTech Model 920 sampler is used to draw the ambient sample through the cartridges.

7.2 Remove the Sep-Pak pair from the transport containers. Uncap and place the cartridges in the metal cartridge holder making sure the short ends of the cartridges point up. Also make sure the primary Sep-Pak cartridge end fits through the 3/16" hole in the septum seal located in the XonTech sampling head receptacle, thus ensuring a leak free system once the cartridge holder is snugged up into the head receptacle.

7.3 Expose the Sep-Pak cartridges to 0.7 LPM ambient air for 24 hours. After the run remove the Sep-Pak cartridge pair, recap the ends and put back in the respective transport containers along with the 920 "printout" for the sampling period.

7.4 The formaldeyde sampling is to be done on the same schedule as that for "Toxics" and therefore the transport container can be placed in the air bag carton for shipment back to the lab.

7.5 Upon receipt at the lab the Sep-Paks are placed in cold storage until desorbtion.

8.0 Sample Analysis

8.1 Sample Desorption

8.1.1 Remove the Sep-Pak cartridges from the culture tube and connect each (outlet or long end during sampling) to a clean syringe.

8.1.2 Place the cartridge/syringe in the syringe rack and backflush the cartridge (gravity feed) with 6 mL of acetonitrile to a 5 mL volumetric flask.

8.1.3 Dilute to the 5 mL mark with ACN. Label the flask with sample ID.

8.2 HPLC Analysis

8.2.1 The operating parameters are as follows:

Column: BondaPak c-18 (3.9 mm x 30 cm) operated @ 30°C

Mobile Phase: Isocratic 40/60 ACN/H_2O

Detector: Model 490 UV/Vis at 360 nM. Sample rate 1 point/sec and 1.00 AUFS.

Flow Rate/Run Time: 1.0 mL/min.; 25 minute run

Retention Time: Formaldehyde ~ 13.5 minutes
Acetaldehyde ~ 19.5 minutes

Injection Volume: 25 uL

8.2.2 Equilibrate the column for 30 minutes before first analysis. Analyze a blank to check for method interferences.

8.2.3 Calibrate the instrument using five standard concentrations each analyzed in triplicate. The results are used to prepare a calibration curve. Linear response is indicated when an r of at least 0.995 for a linear least squares fit of the data is obtained. The retention times for each analyte should agree within 2%.

8.2.4 Check the calibration of the instrument for each run by analyzing a control sample. The concentration given must fall within the UWL and LWL of the control sample value (± 2 S.D.) Plot all results on the method control chart. The day to day response for the various standards should be within 10%.

8.2.5 Prepare a multimethod routine to control the automatic sampler. Run two injections per sample and a control for every ten samples.

9.0 Calculations

$$\text{ug/M}^3 = \frac{(\text{ug/mL})(\text{mL of extract})(1000)}{(\text{minutes sampled})(\text{LPM air flow})}$$

10.0 Method Sensitivity and Precision

Formaldehyde conc.(ug/5mL)	0.4	0.7	1.8	3.6	7.2
Peak Area (M)	55	112	262	566	1098
Std Deviation	4	1	9	36	60
Coeff. of Variation (%)	7.2	0.9	3.4	6.4	5.5

Correlation Coefficient: 0.997

Slope: 0.99996

Intercept: 0.0002 ug/mL

MLD: 0.1 ug/5mL ("X" intercept + 3 S.D. of 0.4 std.)

n=3/conc.

TABLE 1. SENSITIVITY (PPB.V/V) OF SAMPLING/ANALYSIS FOR ALDEHYDES AND KETONES IN AMBIENT AIR USING ADSORBENT CARTRIDGE FOLLOWED BY GRADIENT HIGH PERFORMANCE LIQUID CHROMATOGRAPHY

Sample Volume, L	10	20	30	40	50	60	100	200	300	400	500	1000
Compound	Sensitivity (ppb, v/v) of DNPH/HPLC Method for Carbonyls in Ambient Air											
Formaldehyde	1.45	0.73	0.48	0.36	0.29	0.24	0.15	0.07	0.05	0.04	0.03	0.01
Acetaldehyde	1.36	0.68	0.45	0.34	0.27	0.23	0.14	0.07	0.05	0.03	0.03	0.01
Acrolein	1.29	0.65	0.43	0.32	0.26	0.22	0.13	0.06	0.04	0.03	0.03	0.01
Acetone	1.28	0.64	0.43	0.32	0.26	0.21	0.13	0.06	0.04	0.03	0.03	0.01
Propionaldehyde	1.28	0.64	0.43	0.32	0.26	0.21	0.13	0.06	0.04	0.03	0.03	0.01
Crotonaldehyde	1.22	0.61	0.41	0.31	0.24	0.20	0.12	0.06	0.04	0.03	0.02	0.01
Butyraldehyde	1.21	0.61	0.40	0.30	0.24	0.20	0.12	0.06	0.04	0.03	0.02	0.01
Benzaldehyde	1.07	0.53	0.36	0.27	0.21	0.18	0.11	0.05	0.04	0.03	0.02	0.01
Isovaleraldehyde	1.15	0.57	0.38	0.29	0.23	0.19	0.11	0.06	0.04	0.03	0.02	0.01
Valeraldehyde	1.15	0.57	0.38	0.29	0.23	0.19	0.11	0.06	0.04	0.03	0.02	0.01
o-tolualdehyde	1.02	0.51	0.34	0.25	0.20	0.17	0.10	0.05	0.03	0.03	0.02	0.01
m-tolualdehyde	1.02	0.51	0.34	0.25	0.20	0.17	0.10	0.05	0.03	0.03	0.02	0.01
p-tolualdehyde	1.02	0.51	0.34	0.25	0.20	0.17	0.10	0.05	0.03	0.03	0.02	0.01
Hexanaldehyde	1.09	0.55	0.36	0.27	0.22	0.18	0.11	0.05	0.04	0.03	0.02	0.01
2,5-Dimethylbenzaldehyde	0.97	0.49	0.32	0.24	0.19	0.16	0.10	0.05	0.03	0.02	0.02	0.01

[Note: Ppb values are measured at 1 atm and 25°C; sample cartridge is eluted with 5 mL acetonitrile, and 25 µL are injected onto HPLC column.]

[Note: Maximum sampling flow through a DNPH-coated SEP-PAK is about 1.5 L per minute.]

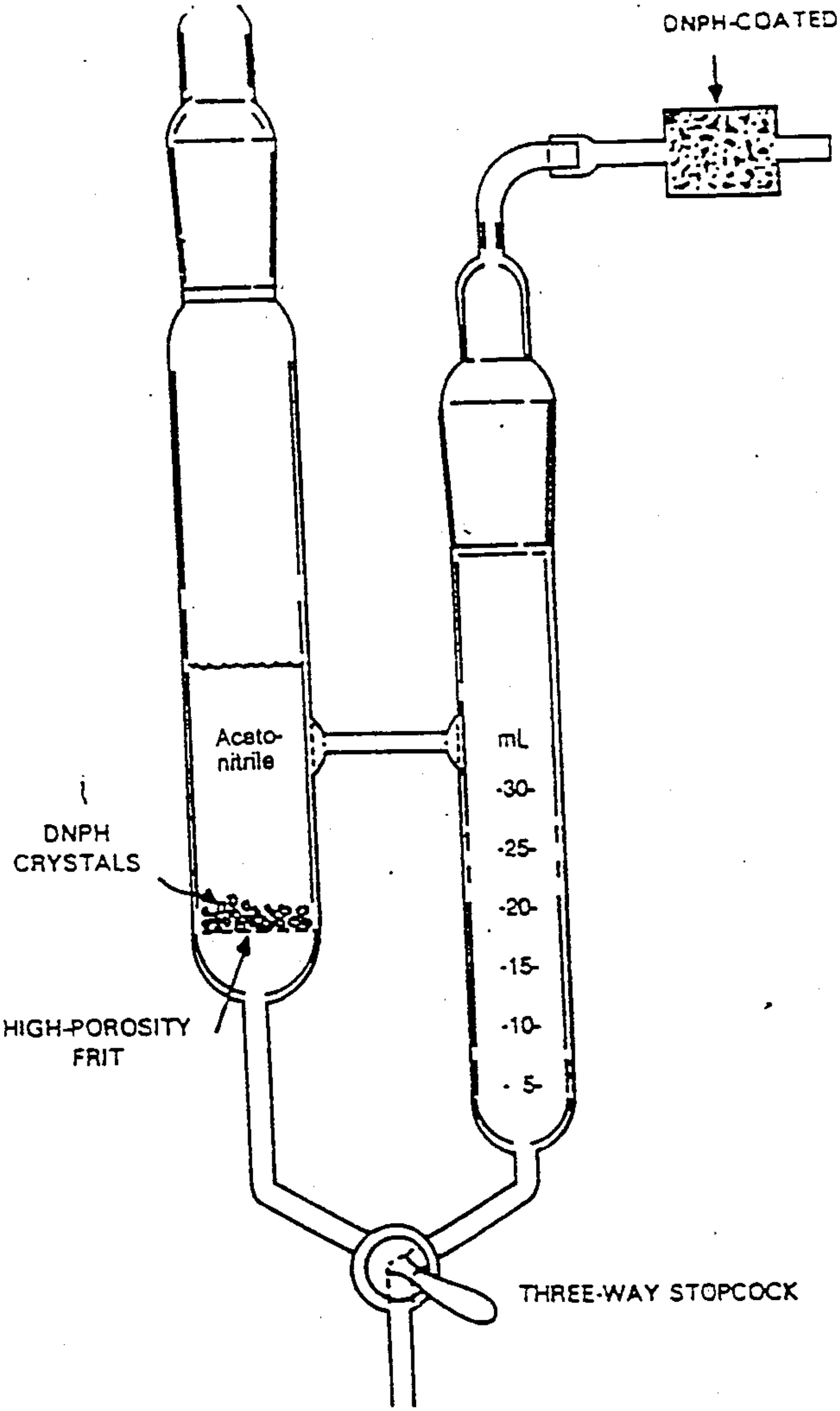


FIGURE 1 SPECIAL GLASS APPARATUS FOR RINSING, STORING, AND DISPENSING SATURATED DNPH STOCK SOLUTION

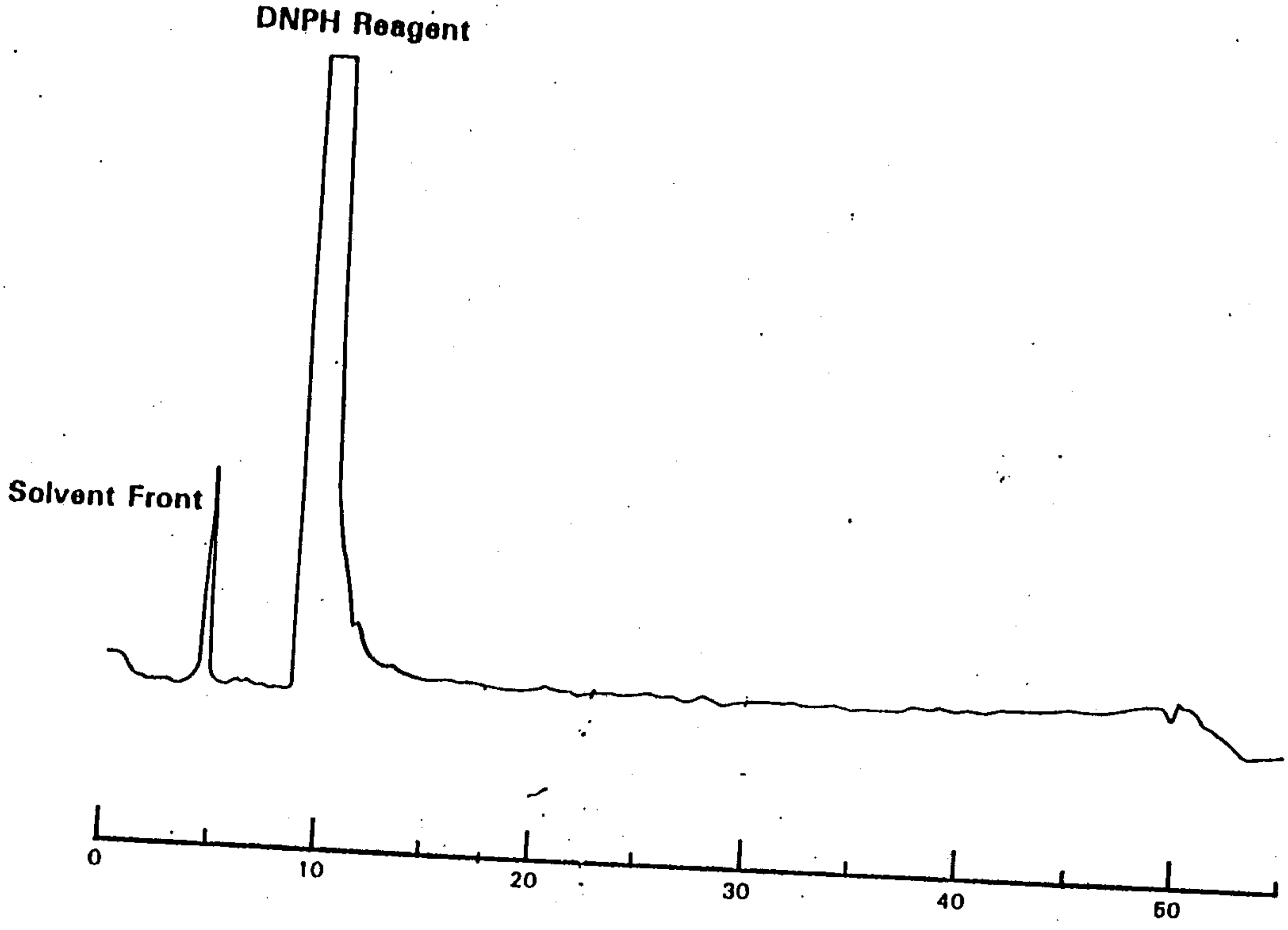
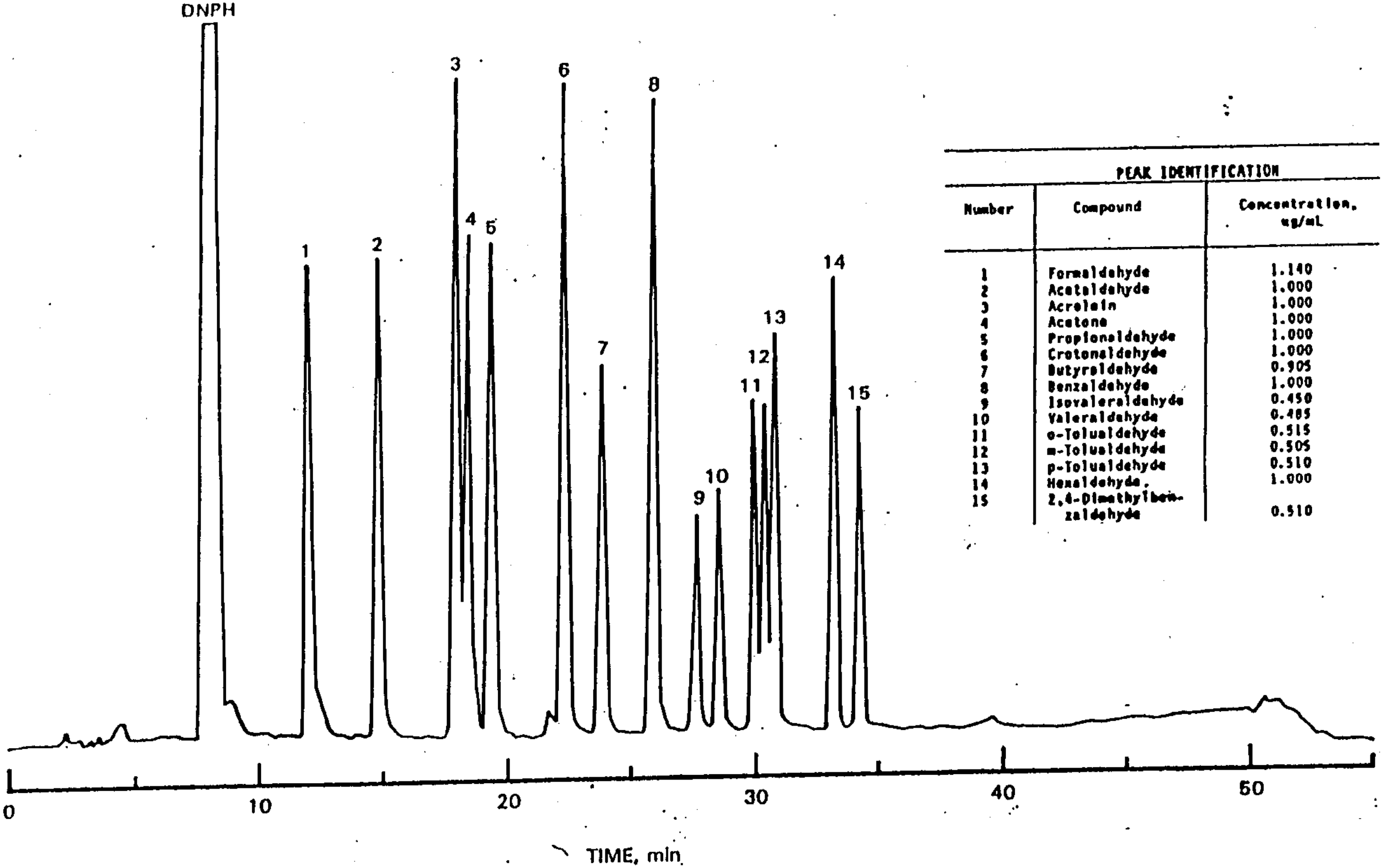


FIGURE 2-. IMPURITY LEVEL OF DNPH
AFTER RECRYSTALLIZATION



PEAK IDENTIFICATION

Number	Compound	Concentration, ug/mL
1	Formaldehyde	1.140
2	Acetaldehyde	1.000
3	Acrolein	1.000
4	Acetone	1.000
5	Propionaldehyde	1.000
6	Crotonaldehyde	1.000
7	Butyraldehyde	0.905
8	Benzaldehyde	1.000
9	Isovaleraldehyde	0.450
10	Valeraldehyde	0.485
11	o-Tolualdehyde	0.515
12	m-Tolualdehyde	0.505
13	p-Tolualdehyde	0.510
14	Hexaldehyde.	1.000
15	2,4-Dimethylbenzaldehyde	0.510

FIGURE 3 CHROMATOGRAPHIC SEPARATION OF DNPH DERIVATIVES OF 15 CARBONYL STANDARDS

APPENDIX D

BOOTSTRAP STATISTICAL METHODS

Statistical Science
1986, Vol. 1, No. 1, 54-77

Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy

B. Efron and R. Tibshirani

Abstract. This is a review of bootstrap methods, concentrating on basic ideas and applications rather than theoretical considerations. It begins with an exposition of the bootstrap estimate of standard error for one-sample situations. Several examples, some involving quite complicated statistical procedures, are given. The bootstrap is then extended to other measures of statistical accuracy such as bias and prediction error, and to complicated data structures such as time series, censored data, and regression models. Several more examples are presented illustrating these ideas. The last third of the paper deals mainly with bootstrap confidence intervals.

Key words: Bootstrap method, estimated standard errors, approximate confidence intervals, nonparametric methods.

1. INTRODUCTION

A typical problem in applied statistics involves the estimation of an unknown parameter θ. The two main questions asked are (1) what estimator $\hat{\theta}$ should be used? (2) Having chosen to use a particular $\hat{\theta}$, how accurate is it as an estimator of θ? The bootstrap is a general methodology for answering the second question. It is a computer-based method, which substitutes considerable amounts of computation in place of theoretical analysis. As we shall see, the bootstrap can routinely answer questions which are far too complicated for traditional statistical analysis. Even for relatively simple problems computer-intensive methods like the bootstrap are an increasingly good data analytic bargain in an era of exponentially declining computational costs.

This paper describes the basis of the bootstrap theory, which is very simple, and gives several examples of its use. Related ideas like the jackknife, the delta method, and Fisher's information bound are also discussed. Most of the proofs and technical details are omitted. These can be found in the references given, particularly Efron (1982a). Some of the discussion here is abridged from Efron and Gong (1983) and also from Efron (1984).

Before beginning the main exposition, we will describe how the bootstrap works in terms of a problem where it is not needed, assessing the accuracy of the sample mean. Suppose that our data consists of a random sample from an unknown probability distribution F on the real line,

$$X_1, X_2, \cdots, X_n \sim F. \tag{1.1}$$

Having observed $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$, we compute the sample mean $\bar{x} = \sum_1^n x_n/n$, and wonder how accurate it is as an estimate of the true mean $\theta = E_F\{X\}$.

If the second central moment of F is $\mu_2(F) \equiv E_F X^2 - (E_F X)^2$, then the standard error $\sigma(F; n, \bar{x})$, that is the standard deviation of $\bar{x}$ for a sample of size n from distribution F, is

$$\sigma(F) = [\mu_2(F)/n]^{1/2}. \tag{1.2}$$

The shortened notation $\sigma(F) \equiv \sigma(F; n, \bar{x})$ is allowable because the sample size n and statistic of interest $\bar{x}$ are known, only F being unknown. The standard error is the traditional measure of $\bar{x}$'s accuracy. Unfortunately, we cannot actually use (1.2) to assess the accuracy of $\bar{x}$, since we do not know $\mu_2(F)$, but we can use the *estimated standard error*

$$\bar{\sigma} = [\hat{\mu}_2/n]^{1/2}, \tag{1.3}$$

where $\hat{\mu}_2 = \sum_{i=1}^n (x_i - \bar{x})^2/(n-1)$, the unbiased estimate of $\mu_2(F)$.

There is a more obvious way to estimate $\sigma(F)$. Let

B. Efron is Professor of Statistics and Biostatistics, and Chairman of the Program in Mathematical and Computational Science at Stanford University. His mailing address is Department of Statistics, Sequoia Hall, Stanford University, Stanford, CA 94305. R. Tibshirani is a Postdoctoral Fellow in the Department of Preventive Medicine and Biostatistics, Faculty of Medicine, University of Toronto, McMurrick Building, Toronto, Ontario, M5S 1A8, Canada.

$\hat F$ indicate the empirical probability distribution.

(1.4) $\hat F$: probability mass $1/n$ on $x_1, x_2, \cdots, x_n$.

Then we can simply replace F by $\hat F$ in (1.2), obtaining

$$\hat\sigma = \sigma(\hat F) = [\mu_2(\hat F)/n]^{1/2}, \quad (1.5)$$

is the estimated standard error for $\bar x$. This is the *bootstrap estimate*. The reason for the name "bootstrap" will be apparent in Section 2, when we evaluate $\sigma(\hat F)$ for statistics more complicated than $\bar x$. Since

$$\hat\mu_2 = \mu_2(\hat F) = \sum_{i=1}^{n} \frac{(x_i - \bar x)^2}{n}, \quad (1.6)$$

$\hat\sigma$ is not quite the same as $\bar\sigma$, but the difference is too small to be important in most applications.

Of course we do not really need an alternative formula to (1.3) in this case. The trouble begins when we want a standard error for estimators more complicated than $\bar x$, for example, a median or a correlation or a slope coefficient from a robust regression. In most cases there is no equivalent to formula (1.2), which expresses the standard error $\sigma(F)$ as a simple function of the sampling distribution F. As a result, formulas like (1.3) do not exist for most statistics.

This is where the computer comes in. It turns out that we can always numerically evaluate the bootstrap estimate $\hat\sigma = \sigma(\hat F)$, without knowing a simple expression for $\sigma(F)$. The evaluation of $\hat\sigma$ is a straightforward Monte Carlo exercise described in the next section. In a good computing environment, as described in the remarks in Section 2, the bootstrap effectively gives the statistician a simple formula like (1.3) for any statistic, no matter how complicated.

Standard errors are crude but useful measures of statistical accuracy. They are frequently used to give approximate confidence intervals for an unknown parameter θ

$$\theta \in \hat\theta \pm \hat\sigma z^{(\alpha)}, \quad (1.7)$$

where $z^{(\alpha)}$ is the $100 \cdot \alpha$ percentile point of a standard normal variate, e.g., $z^{(.95)} = 1.645$. Interval (1.7) is sometimes good, and sometimes not so good. Sections 7 and 8 discuss a more sophisticated use of the bootstrap, which gives better approximate confidence intervals than (1.7).

The standard interval (1.7) is based on taking literally the large sample normal approximation $(\hat\theta - \theta)/\hat\sigma \sim N(0, 1)$. Applied statisticians use a variety of tricks to improve this approximation. For instance if θ is the correlation coefficient and $\hat\theta$ the sample correlation, then the transformation $\phi = \tanh^{-1}(\theta)$, $\hat\phi = \tanh^{-1}(\hat\theta)$ greatly improves the normal approximation, at least in those cases where the underlying sampling distribution is bivariate normal. The correct tactic then is to transform, compute the interval (1.7) for ϕ, and transform this interval back to the θ scale.

We will see that bootstrap confidence intervals can automatically incorporate tricks like this, without requiring the data analyst to produce special techniques, like the $\tanh^{-1}$ transformation, for each new situation. An important theme of what follows is the substitution of raw computing power for theoretical analysis. This is not an argument against theory, of course, only against unnecessary theory. Most common statistical methods were developed in the 1920s and 1930s, when computation was slow and expensive. Now that computation is fast and cheap we can hope for and expect changes in statistical methodology. This paper discusses one such potential change, Efron (1979b) discusses several others.

2. THE BOOTSTRAP ESTIMATE OF STANDARD ERROR

This section presents a more careful description of the bootstrap estimate of standard error. For now we will assume that the observed data $\mathbf{y} = (x_1, x_2, \cdots, x_n)$ consists of independent and identically distributed (iid) observations $X_1, X_2, \cdots, X_n \sim_{\text{iid}} F$, as in (1.1). Here F represents an unknown probability distribution on $\mathscr{X}$, the common sample space of the observations. We have a statistic of interest, say $\hat\theta(\mathbf{y})$, to which we wish to assign an estimated standard error.

Fig. 1 shows an example. The sample space $\mathscr{X}$ is $\mathbb{R}^{2+}$, the positive quadrant of the plane. We have observed $n = 15$ bivariate data points, each corresponding to an American law school. Each point x_i consists of two summary statistics for the 1973 entering class at law school i

$$x_i = (\text{LSAT}_i, \text{GPA}_i); \quad (2.1)$$

LSAT$_i$ is the class' average score on a nationwide exam called "LSAT"; GPA$_i$ is the class' average undergraduate grades. The observed Pearson correlation

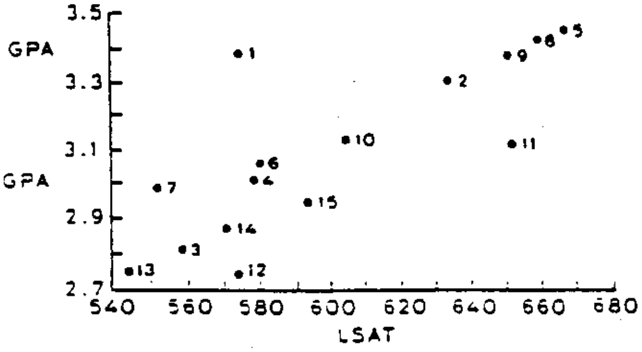


FIG. 1. *The law school data (Efron, 1979b). The data points, beginning with School 1, are* (576, 3.39), (635, 3.30), (558, 2.81), (578, 3.03), (666, 3.44), (580, 3.07), (555, 3.00), (661, 3.43), (651, 3.36), (605, 3.13), (653, 3.12), (575, 2.74), (545, 2.76), (572, 2.88), (594, 2.96).

coefficient for these 15 points is $\hat\theta = .776$. We wish to assign a standard error to this estimate.

Let $\sigma(F)$ indicate the standard error of $\hat\theta$, as a function of the unknown sampling distribution F,

$$\sigma(F) = [\mathrm{Var}_F\{\hat\theta(\mathbf{y})\}]^{1/2}. \tag{2.2}$$

Of cou[illegible] a function of the sample size n and the form of the statistic $\hat\theta(\mathbf{y})$, but since both of these are known they need not be indicated in the notat[illegible] bootstrap estimate of standard error is

$$\hat\sigma = \sigma(\hat F), \tag{2.3}$$

wher[illegible] the empirical distribution (1.4), putting [illegible] data point x_i. In the [illegible]ution putting mass $\frac{1}{15}$ on each point in Fig. 1, and $\hat\sigma$ is the standard deviation of the correlation coefficient for 15 iid points draw[illegible].

In most cases, including that of the correlation coefficient, there is no simple expression for the function $\sigma(F)$ in (2.2). Nevertheless, it is easy to numerically evaluate $\hat\sigma = \sigma(\hat F)$ by means of a Monte Carlo algorithm, which depends on the following notation: $\mathbf{y}^* = ($[illegible]$, x_n^*)$ indicates n independent draws from $\hat F$, called a *bootstrap sample*. Because $\hat F$ is the empirical distribution of the data, a bootstrap sample turn[illegible] same as a random sample of size n d[illegible] *[illegible]placement* from the actual sample $\{x_1,$ [illegible] $x_n\}$.

The Monte Carlo algorithm proceeds in three steps: (i) using a random number generator, independently draw a large number of bootstrap samples, say $\mathbf{y}^*(1)$, $\mathbf{y}^*(2)$, [illegible] $\mathbf{y}^*(B)$; (ii) for each bootstrap sample $\mathbf{y}^*(b)$, evaluate the statistic of interest, say $\hat\theta^*(b) = \hat\theta(\mathbf{y}^*(b))$, $b = 1, 2,$ [illegible] B; and (iii) calculate the sample standard de[illegible] values

$$\hat\sigma_B = \left(\frac{\sum_{b=1}^{B}[\hat\theta^*(b) - \hat\theta^*(\cdot)]^2}{B-1}\right)^{1/2}, \qquad \hat\theta^*(\cdot) = \frac{\sum_{b=1}^{B}\hat\theta^*(b)}{B}. \tag{2.4}$$

It is easy to see that as $B \to \infty$, $\hat\sigma_B$ will approach $\hat\sigma$ [illegible] bootstrap estimate of standard error. All we are doing is evaluating a standard deviation by Monte Carlo sampling. Later, in Section 9, we will discuss how large B need be taken. For most [illegible] range 50 to 200 is quite adequate. In what follows we will usually ignore the difference between $\hat\sigma_B$ and $\hat\sigma$, calling both simply "$\hat\sigma$."

Why [illegible] bootstrap sample taken with the same sample size n as the original data set? Remember that $\sigma(F)$ is actually $\sigma(F, n, \hat\theta)$, the standard error for the statistic $\hat\theta(\cdot)$ based on a random sample of size n from [illegible] the bootstrap estimate [illegible] is actually $\sigma(F, n, \hat\theta)$ evaluated at $F = \hat F$. The Monte Carlo algorithm will not converge to $\hat\sigma$ if the bootstrap sample size differs from the true n. Bickel and Freedman (1981) show how to correct the algorithm to give $\hat\sigma$ if in fact the bootstrap sample size is taken different than n, but so far there does not seem to be any practical advantage to be gained in this way.

Fig. 2 shows the histogram of $B = 1000$ bootstrap replications of the correlation coefficient from the law school data. For convenient reference the abscissa is plotted in terms of $\hat\theta^* - \hat\theta = \hat\theta^* - .776$. Formula (2.4) gives $\hat\sigma = .127$ as the bootstrap estimate of standard error. This can be compared with the usual normal theory estimate of standard error for $\hat\theta$,

$$\hat\sigma_{\mathrm{NORM}} = (1 - \hat\theta^2)/(n-3)^{1/2} = .115, \tag{2.5}$$

[Johnson and Kotz (1970, p. 229)].

Remark. The Monte Carlo algorithm leading to $\hat\sigma_B$ (2.4) is simple to program. On the Stanford version of the statistical computing language S, Professor Arthur Owen has introduced a single command which bootstraps any statistic in the S catalog. For instance the bootstrap results in Fig. 2 are obtained simply by typing

```
tboot(lawdata, correlation, B = 1000).
```

The execution time is about a factor of B greater than that for the original computation.

There is another way to describe the bootstrap standard error: $\hat F$ is the nonparametric maximum likelihood estimate (MLE) of the unknown distribution F (Kiefer and Wolfowitz, 1956). This means that the bootstrap estimate $\hat\sigma = \sigma(\hat F)$ is the nonparametric MLE of $\sigma(F)$, the true standard error.

In fact there is nothing which says that the bootstrap must be carried out nonparametrically. Suppose for instance that in the law school example we believe the true sampling distribution F must be bivariate normal. Then we could estimate F with its *parametric* MLE $\hat F_{\mathrm{NORM}}$, the bivariate normal distribution having the same mean vector and covariance matrix as the

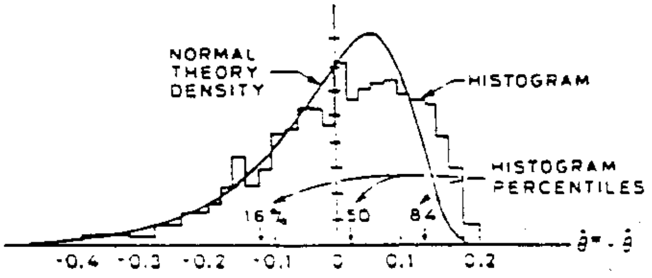


Fig. 2. *Histogram of $B = 1000$ bootstrap replications of $\hat\theta^*$ for the law school data. The normal theory density curve has a similar shape, but falls off more quickly at the upper tail.*

data. The bootstrap samples at step (i) of the algorithm could then be drawn from $\hat{F}_{\text{NORM}}$ instead of $\hat{F}$, and steps (ii) and (iii) carried out as before.

The smooth curve in Fig. 2 shows the results of carrying out this "normal theory bootstrap" on the law school data. Actually there is no need to do the bootstrap sampling in this case, because of Fisher's formula for the sampling density of a correlation coefficient in the bivariate normal situation (see Chapter 32 of Johnson and Kotz, 1970). This density can be thought of as the bootstrap distribution for $B = \infty$. Expression (2.5) is a close approximation to $\hat{\sigma}_{\text{NORM}} = \sigma(\hat{F}_{\text{NORM}})$, the parametric bootstrap estimate of standard [illegible]

[illegible] of the bootstrap, [illegible] worth remembering that all of the usual formulas for estimating standard errors, like $\hat{\mathscr{J}}^{-1/2}$ where [illegible] is the observed Fisher information, are essentially bootstrap estimates carried out in a parametric framework. This point is carefully explained in Section 5 of Efron (1982c). The straightforward nonparametric algorithm (i)–(iii) has the virtues of avoiding all parametric assumptions, all approximations (such as those involved with the Fisher information expression for the standard error of an MLE), and in fact all analytic difficulties of any kind. The data analyst is free to obtain standard errors for enormously complicated estimators, subject only to the constraints of computer time. Sections 3 and 6 discuss some interesting applied problems which are far too complicated for standard analyses.

How well does the bootstrap work? Table 1 shows the answer in one situation. Here $\mathscr{X}$ is the real line, $n = 15$, and the statistic $\hat{\theta}$ of interest is the 25% trimmed mean. If the true sampling distribution F is $N(0, 1)$, then the true standard error is $\sigma(F) = .286$. The bootstrap estimate $\hat{\sigma}$ is nearly unbiased, averaging .287 in a large sampling experiment. The standard deviation of the bootstrap estimate $\hat{\sigma}$ is itself .071 in this case, with coefficient of variation .071/.287 = .25. (Notice that there are two levels of Monte Carlo involved in Table 1: first drawing the actual samples $\mathbf{y} = (x_1, x_2, \cdots, x_{15})$ from F, and then drawing bootstrap samples $(x_1^*, x_2^*, \cdots, x_{15}^*)$ with $\mathbf{y}$ held fixed. The bootstrap samples evaluate $\hat{\sigma}$ for a fixed value of $\mathbf{y}$. The standard deviation .071 refers to the variability of $\hat{\sigma}$ due to the random choice of $\mathbf{y}$.)

The jackknife, another common method of assigning nonparametric standard errors, is discussed in Section 10. The jackknife estimate $\hat{\sigma}_J$ is also nearly unbiased for $\sigma(F)$, but has higher coefficient of variation (CV). The minimum possible CV for a scale-invariant estimate of $\sigma(F)$, assuming full knowledge of the parametric model, is shown in brackets. The nonparametric bootstrap is seen to be moderately efficient in both cases considered in Table 1.

Table 2 returns to the case of $\hat{\theta}$ the correlation coefficient. Instead of real data we have a sampling experiment in which the true F is bivariate normal, true correlation $\theta = .50$, sample size $n = 14$. Table 2 is abstracted from a larger table in Efron (1981b), in

TABLE 1

A sampling experiment comparing the bootstrap and jackknife estimates of standard error for the 25% trimmed mean, sample size $n = 15$

	F standard normal			F negative exponential		
	Ave	SD	CV	Ave	SD	CV
Bootstrap $\hat{\sigma}$ $(B = 200)$	.287	.071	.25	.242	.078	.32
Jackknife	.280	.084	.30	.224	.085	.38
True [illegible]	[illegible]		(.19)	.232		(.27)

TABLE 2

Estimates of standard error for the correlation coefficient $\hat{\theta}$ and for $\hat{\phi} = \tanh^{-1}\hat{\theta}$; sample size $n = 14$, distribution F bivariate normal with true correlation $\rho = .5$ (from a larger table in Efron, 1981b)

	Summary statistics for 200 trials							
	Standard error estimates for $\hat{\theta}$				Standard error estimates for $\hat{\phi}$			
	Ave	SD	CV	$\sqrt{\text{MSE}}$	Ave	SD	CV	$\sqrt{\text{MSE}}$
1. Bootstrap $B = 128$	.206	.066	.32	.067	.301	.065	.22	.065
2. Bootstrap $B = 512$	.206	.063	.31	.064	.301	.062	.21	.062
3. Normal smoothed bootstrap $B = 128$	.200	.060	.30	.063	.296	.041	.14	.041
4. Uniform smoothed bootstrap $B = 128$	.205	.061	.30	.062	.298	.058	.19	.058
5. Uniform smoothed bootstrap $B = 512$	.205	.059	.29	.060	.296	.052	.18	.052
6. Jackknife	.223	.085	.38	.085	.314	.090	.29	.091
7. Delta method (Infinitesimal jackknife)	.175	.058	.33	.072	.244	.052	.21	.076
8. Normal theory	.217	.056	.26	.056	.302	0	0	.003
True standard error	.218				.299			

which some of the methods for estimating a standard error required the sample size to be even.

The left side of Table 2 refers to $\hat{\theta}$, while the right side refers to $\hat{\phi} = \tanh^{-1}(\hat{\theta}) = .5 \log(1 + \hat{\theta})/(1 - \hat{\theta})$. For each estimator of standard error, the root mean squared error of estimation $[E(\hat{\sigma} - \sigma)^2]^{1/2}$ is given in the column headed $\sqrt{\text{MSE}}$.

The bootstrap was run with $B = 128$ and also with $B = 512$, the latter value yielding only slightly better estimates in accordance with the results of Section 9. Further increasing B would be pointless. It can be shown that $B = \infty$ gives $\sqrt{\text{MSE}} = .063$ for $\hat{\theta}$, only .001 less than $B = 512$. The normal theory estimate (2.5), which we know to be ideal for this sampling experiment, has $\sqrt{\text{MSE}} = .056$.

We can compromise between the totally nonparametric bootstrap estimate $\hat{\sigma}$ and the totally parametric bootstrap estimate $\hat{\sigma}_{\text{NORM}}$. This is done in lines 3, 4, and 5 of Table 2. Let $\hat{\Sigma} = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})'/n$ be the sample covariance matrix of the observed data. The *normal smoothed bootstrap* draws the bootstrap sample from $\hat{F} \oplus N_2(0, .25\hat{\Sigma})$, $\oplus$ indicating convolution. This amounts to estimating F by an equal mixture of the n distributions $N_2(x_i, .25\hat{\Sigma})$, that is by a normal window estimate. Each point x_i^* in a smoothed bootstrap sample is the sum of a randomly selected original data point x_j, plus an independent bivariate normal point $z_j \sim N_2(0, .25\hat{\Sigma})$. Smoothing makes little difference on the left side of the table, but is spectacularly effective in the $\hat{\phi}$ case. The latter result is suspect since the true sampling distribution is bivariate normal, and the function $\hat{\phi} = \tanh^{-1}\hat{\theta}$ is specifically chosen to have nearly constant standard error in the bivariate normal family. The *uniform smoothed bootstrap* samples from $\hat{F} \oplus \mathfrak{U}(0, .25\hat{\Sigma})$, where $\mathfrak{U}(0, .25\hat{\Sigma})$ is the uniform distribution on a rhombus selected so $\mathfrak{U}$ has mean vector 0 and covariance matrix $.25\hat{\Sigma}$. It yields moderate reductions in $\sqrt{\text{MSE}}$ for both sides of the table.

Line 6 of Table 2 refers to the *delta method*, which is the most common method of assigning nonparametric standard error. Surprisingly enough, it is badly biased downward on both sides of the table. The delta method, also known as the method of statistical differentials, the Taylor series method, and the infinitesimal jackknife, is discussed in Section 10.

3. EXAMPLES

Example 1. Cox's Proportional Hazards Model

In this section we apply bootstrap standard error estimation to some complicated statistics.

The data for this example come from a study of leukemia remission times in mice, taken from Cox (1972). They consist of measurements of remission time (y) in weeks for two groups, treatment ($x = 0$) and control ($x = 1$), and a 0–1 variable (δ_i) indicating whether or not the remission time is censored (0) or complete (1). There are 21 mice in each group.

The standard regression model for censored data is Cox's proportional hazards model (Cox, 1972). It assumes that the hazard function $h(t \mid x)$, the probability of going into remission in next instant given no remission up to time t for a mouse with covariate x, is of the form

$$h(t \mid x) = h_0(t)e^{\beta x}. \tag{3.1}$$

Here $h_0(t)$ is an arbitrary unspecified function. Since x here is a group indicator, this means simply that the hazard for the control group is e^{β} times the hazard for the treatment group. The regression parameter β is estimated independently of $h_0(t)$ through maximization of the so called "partial likelihood"

$$\text{PL} = \prod_{i \in D} \frac{e^{\beta x_i}}{\sum_{j \in R_i} e^{\beta x_j}}, \tag{3.2}$$

where D is the set of indices of the failure times and R_i is the set of indices of those at risk at time y_i. This maximization requires an iterative computer search.

The estimate $\hat{\beta}$ for these data turns out to be 1.51. Taken literally, this says that the hazard rate is $e^{1.51} = 4.33$ times higher in the control group than in the treatment group, so the treatment is very effective. What is the standard error of $\hat{\beta}$? The usual asymptotic maximum likelihood theory, one over the square root of the observed Fisher information, gives an estimate of .41. Despite the complicated nature of the estimation procedure, we can also estimate the standard error using the bootstrap. We sample with replacement from the triples $\{(y_1, x_1, \delta_1), \cdots, (y_{42}, x_{42}, \delta_{42})\}$. For each bootstrap sample $\{(y_1^*, x_1^*, \delta_1^*), \cdots, (y_{42}^*, x_{42}^*, \delta_{42}^*)\}$ we form the partial likelihood and numerically maximize it to produce the bootstrap estimate $\hat{\beta}^*$. A histogram of 1000 bootstrap values is shown in Fig. 3.

The bootstrap estimate of the standard error of $\hat{\beta}$ based on these 1000 numbers is .42. Although the bootstrap and standard estimates agree, it is interesting to note that the bootstrap distribution is skewed to the right. This leads us to ask: is there other information that we can extract from the bootstrap distribution other than a standard error estimate? The answer is yes—in particular, the bootstrap distribution can be used to form a confidence interval for β, as we will see in Section 9. The shape of the bootstrap distributiion will help determine the shape of the confidence interval.

In this example our resampling unit was the triple (y_i, x_i, δ_i), and we ignored the unique elements of the problem, i.e., the censoring, and the particular model being used. In fact, there are other ways to bootstrap

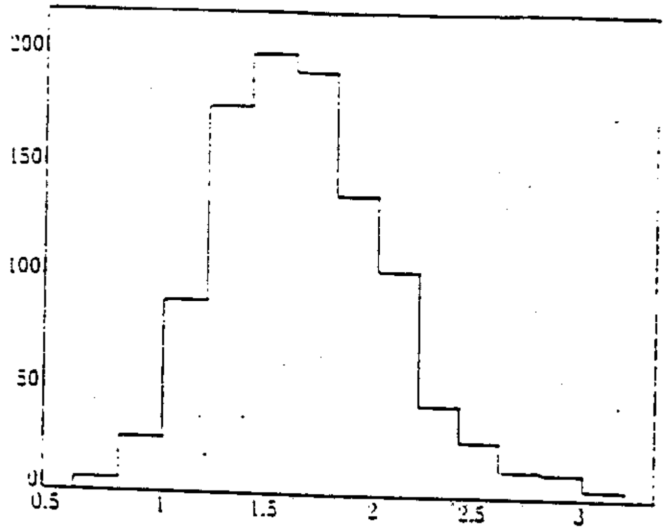


FIG. 3. *Histogram of* 1000 *bootstrap replications for the mouse leukemia data.*

this problem. We will see this when we discuss bootstrapping censored data in Section 5.

Example 2: Linear and Projection Pursuit Regression

We illustrate an application of the bootstrap to standard linear least squares regression as well as to a nonparametric regression technique.

Consider the standard regression setup. We have n observations on a response Y and covariates $(X_1, X_2, \cdots, X_p)$. Denote the ith observed vector of covariates by $x_i = (x_{i1}, x_{i2}, \cdots, x_{ip})'$. The usual linear regression model assumes

$$E(Y_i) = \alpha + \sum_{j=1}^{p} \beta_j x_{ij}. \tag{3.3}$$

Friedman and Stuetzle (1981) introduced a more general model, the *projection pursuit* regression model

$$E(Y_i) = \sum_{j=1}^{m} s_j(a_j \cdot x_i). \tag{3.4}$$

The p vectors a_j are unit vectors ("directions"), and the functions $s_j(\cdot)$ are unspecified.

Estimation of $\{a_1, s_1(\cdot)\}, \cdots, \{a_m, s_m(\cdot)\}$ is performed in a forward stepwise manner as follows. Consider $\{a_1, s_1(\cdot)\}$. Given a direction a_1, $s_1(\cdot)$ is estimated by a nonparametric smoother (e.g., running mean) of y on $a_1 \cdot x$. The projection pursuit regression algorithm searches over all unit directions to find the direction $\hat a_1$ and associated function $\hat s_1(\cdot)$ that minimize $\sum_1^n (y_i - \hat s_1(\hat a \cdot x_i))^2$. Then residuals are taken and the next direction and function are determined. This process is continued until no additional term significantly reduces the residual sum of squares.

Notice the relation of the projection pursuit regression model to the standard linear regression model. When the function $s_1(\cdot)$ is forced to be linear and is estimated by the usual least squares method, a one-term projection pursuit model is exactly the same as the standard linear regression model. That is to say, the fitted model $\hat s_1(\hat a_1 \cdot x_i)$ exactly equals the least squares fit $\hat\alpha + \sum_{j=1}^{p} \hat\beta_j x_{ij}$. This is because the least squares fit, by definition, finds the best direction and the best linear function of that direction. Note also that adding another linear term $\hat s_2(\hat a_2 \cdot x_i)$ would not change the fitted model since the sum of two linear functions is another linear function.

Hastie and Tibshirani (1984) applied the bootstrap to the linear and projection pursuit regression models to assess the variability of the coefficients in each. The data they considered are taken from Breiman and Friedman (1985). The response Y is Upland atmospheric ozone concentration (ppm); the covariates X_1 = Sandburg Air Force base temperature (C°), X_2 = inversion base height (ft), X_3 = Daggot pressure gradient (mm Hg), X_4 = visibility (miles), and X_5 = day of the year. There are 330 observations. The number of terms (m) in the model (3.4) is taken to be two. The projection pursuit algorithm chose directions $\hat a_1 = (.80, -.38, .37, -.24, -.14)'$ and $\hat a_2 = (.07, .16, .04, -.05, -.98)'$. These directions consist mostly of Sandburg Air Force temperature and day of the year, respec-

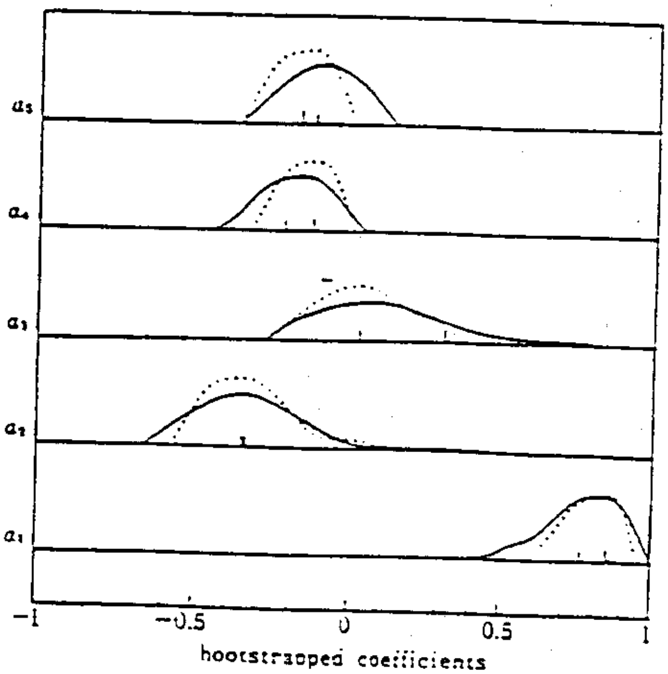


FIG. 4. *Smoothed histograms of the bootstrapped coefficients for the first term in the projection pursuit regression model. Solid histograms are for the usual projection pursuit model; the dotted histograms are for linear* $s_1(\cdot)$.

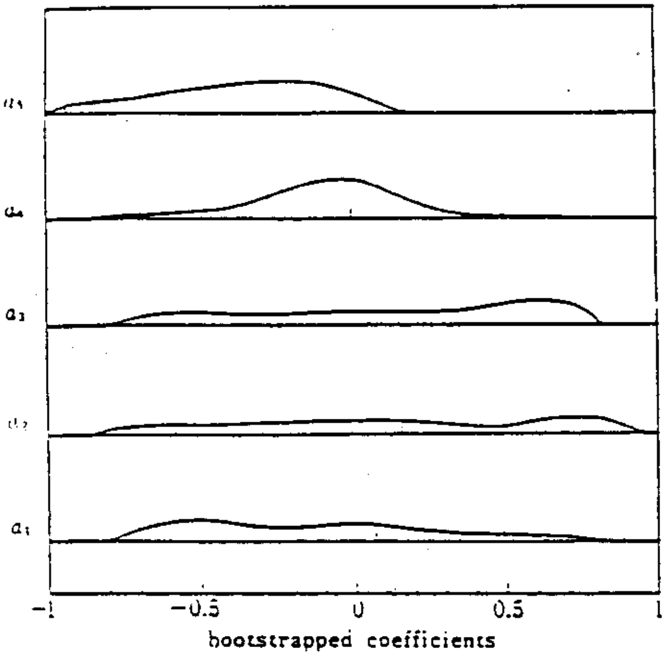


FIG. 5. *Smoothed histograms of the bootstrapped coefficients for the second term in the projection pursuit model.*

tively. (We do not show graphs of the estimated functions $\hat{s}_1(\cdot)$ and $\hat{s}_2(\cdot)$ although in a full analysis of the data they would also be of interest.) Forcing $\hat{s}_1(\cdot)$ to be linear results in the direction $\hat{\alpha}_1 = (.90, -.37, .03, -.14, -.19)'$. These are just the usual least squares estimates $\hat{\beta}_1, \cdots, \hat{\beta}_p$ scaled so that $\sum_1^p \beta_j^2 = 1$.

To assess the variability of the directions, a bootstrap sample is drawn with replacement from $(y_1, x_{11}, \cdots, x_{15}), \cdots, (y_{330}, x_{3301}, \cdots, x_{3305})$ and the projection pursuit algorithm is applied. Figs. 4 and 5 show histograms of the directions $\hat{\alpha}_1^*$ and $\hat{\alpha}_2^*$ for 200 bootstrap replications. Also shown in Fig. 4 (broken histogram) are the bootstrap replications of $\hat{\alpha}_1$ with $\hat{s}_1(\cdot)$ forced to be linear.

The first direction of the projection pursuit model is quite stable and only slightly more variable than the corresponding linear regression direction. But the second direction is extremely unstable! It is clearly unwise to put any faith in the second direction of the original projection pursuit model.

Example 3: Cox's Model and Local Likelihood Estimation

In this example, we return to Cox's proportional hazards model described in Example 1, but with a few added twists.

The data that we will discuss come from the Stanford heart transplant program and are given in Miller and Halpern (1982). The response y is survival time in weeks after a heart transplant, the covariate x is age at transplant, and the 0–1 variable δ indicates whether the survival time is censored (0) or complete (1). There are measurements on 157 patients. A proportional hazards model was fit to these data, with a quadratic term, i.e. $h(t \mid x) = h_0(t)e^{\beta_1 x + \beta_2 x^2}$. Both $\hat{\beta}_1$ and $\hat{\beta}_2$ are highly significant: the broken curve in Fig. 6 is $\hat{\beta}_1 x + \hat{\beta}_2 x^2$ as a function of x.

For comparison, Fig. 6 shows (solid line) another estimate. This was computed using *local likelihood estimation* (Tibshirani and Hastie, 1984). Given a general proportional hazards model of the form $h(t \mid x) = h_0(t)e^{s(x)}$, the local likelihood technique assumes nothing about the parametric form of $s(x)$; instead it estimates $s(x)$ nonparametrically using a kind of local averaging. The algorithm is very computationally intensive, and standard maximum likelihood theory cannot be applied.

A comparison of the two functions reveals an important qualitative difference: the parametric estimate suggests that the hazard decreases sharply up to age 34, then rises; the local likelihood estimate stays approximately constant up to age 45 then rises. Has the forced fitting of a quadratic function produced a misleading result? To answer this question, we can bootstrap the local likelihood estimate. We sample with replacement from the triples $\{(y_1, x_1, \delta_1) \cdots (y_{157}, x_{157}, \delta_{157})\}$ and apply the local likelihood algorithm to each bootstrap sample. Fig. 7 shows estimated curves from 20 bootstrap samples.

Some of the curves are flat up to age 45, others are decreasing. Hence the original local likelihood estimate is highly variable in this region and on the basis of these data we cannot determine the true behavior of the function there. A look back at the original data shows that while half of the patients were under 45, only 13% of the patients were under 30. Fig. 7 also shows that the estimate is stable near the middle ages but unstable for the older patients.

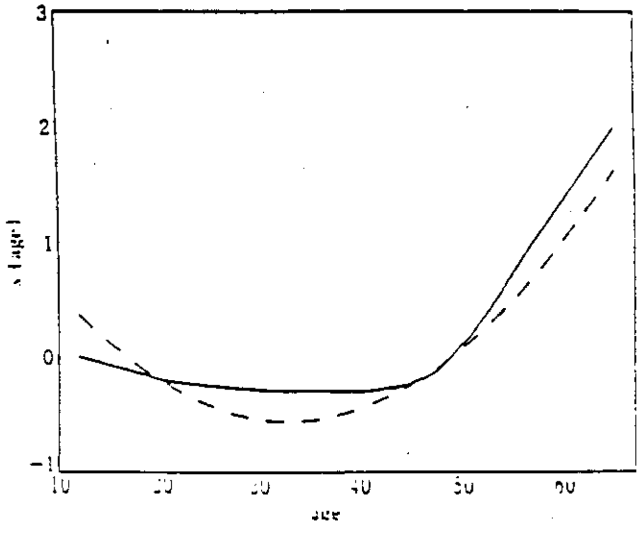


FIG. 6. *Estimates of log relative risk for the Stanford heart transplant data. Broken curve: parametric estimate. Solid curve: local likelihood estimate.*

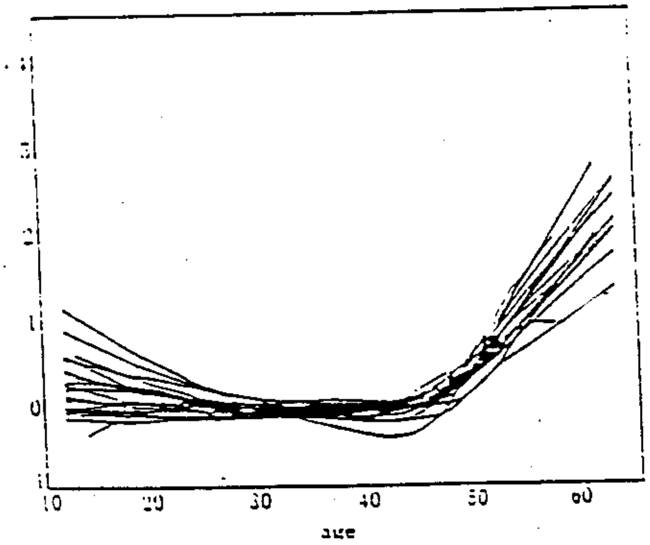


FIG. 7. 20 bootstraps of the local likelihood estimate for the Stanford heart transplant data.

4. OTHER MEASURES OF STATISTICAL ERROR

So far we have discussed statistical error, or accuracy, in terms of the standard error. It is easy to assess other measures of statistical error, such as bias or prediction error, using the bootstrap.

Consider the estimation of bias. For a given statistic $\hat\theta(\mathbf{y})$, and a given parameter $\mu(F)$, let

$$R(\mathbf{y}, F) = \hat\theta(\mathbf{y}) - \mu(F). \tag{4.1}$$

(It will help keep our notation clear to call the parameter of interest μ rather than θ.) For example, μ might be the mean of the distribution F, assuming the sample space $\mathcal{X}$ is the real line, and $\hat\theta$ the 25% trimmed mean. The bias of $\hat\theta$ for estimating μ is

$$\beta(F) = E_F R(\mathbf{y}, F) = E_F\{\hat\theta(\mathbf{y})\} - \mu(F). \tag{4.2}$$

The notation E_F indicates expectation with respect to the probability mechanism appropriate to F, in this case $\mathbf{y} = (x_1, x_2, \cdots, x_n)$ a random sample from F.

The bootstrap estimate of bias is

$$\hat\beta = \beta(\hat F) = E_{\hat F} R(\mathbf{y}^*, \hat F) = E_{\hat F}\{\hat\theta(\mathbf{y}^*)\} - \mu(\hat F). \tag{4.3}$$

As in Section 2, $\mathbf{y}^*$ denotes a random sample $(x_1^*, x_2^*, \cdots, x_n^*)$ from $\hat F$, i.e., a bootstrap sample. To numerically evaluate $\hat\beta$, all we do is change step (iii) of the bootstrap algorithm in Section 2 to

$$\begin{aligned} \hat\beta_B &= \frac{1}{B}\sum_{b=1}^{B} R(\mathbf{y}^*(b), \hat F) \\ &= \frac{\sum_{b=1}^{B} \hat\theta^*(b)}{B} - \mu(\hat F) \\ &= \hat\theta^*(\cdot) - \mu(\hat F). \end{aligned} \tag{4.4}$$

As $B \to \infty$, $\hat\beta_B$ goes to $\hat\beta$ (4.3).

TABLE 3

βHCG blood serum levels for 54 patients having metasticized breast cancer in ascending order

0.1, 0.1, 0.2, 0.4, 0.4, 0.6, 0.8, 0.8, 0.9, 0.9, 1.2, 1.3, 1.4, 1.5, 1.6,
1.6, 1.7, 1.7, 1.7, 1.8, 2.0, 2.0, 2.2, 2.2, 2.2, 2.3, 2.3, 2.4, 2.4, 2.4,
2.4, 2.4, 2.4, 2.5, 2.5, 2.5, 2.7, 2.7, 2.8, 2.9, 2.9, 2.9, 3.0, 3.1, 3.1,
3.2, 3.2, 3.3, 3.3, 3.5, 4.4, 4.5, 6.4, 9.4

As an example consider the blood serum data of Table 3. Suppose we wish to estimate the true mean $\mu = E_F\{X\}$ of this population using $\hat\theta$, the 25% trimmed mean. We calculate $\hat\mu = \mu(\hat F) = 2.32$, the sample mean of the 54 observations, and $\hat\theta = 2.24$, the trimmed mean. The trimmed mean is lower because it discounts the effect of the large observations 6.4 and 9.4. It looks like the trimmed mean might be more robust for this type of data, and as a matter of fact a bootstrap analysis, $B = 1000$, gave estimated standard error $\hat\sigma = .16$ for $\hat\theta$, compared to .21 for the sample mean. But what about bias?

The same 1000 bootstrap replications which gave $\hat\sigma = .16$ also gave $\hat\theta^*(\cdot) = 2.29$, so

$$\hat\beta = 2.29 - 2.32 = -0.03 \tag{4.5}$$

according to (4.4). (The estimated standard deviation of $\hat\beta_B - \hat\beta$ due to the limitations of having $B = 1000$ bootstraps is only 0.005 in this case, so we can ignore the difference between $\hat\beta_B$ and $\hat\beta$.) Whether or not a bias of magnitude −0.03 is too large depends on the context of the problem. If we attempt to remove the bias by subtraction, we get $\hat\theta - \hat\beta = 2.24 - (-0.03) = 2.27$. Removing bias in this way is frequently a bad idea (see Hinkley, 1978), but at least the bootstrap analysis has given us a reasonable picture of the bias and standard error of $\hat\theta$.

Here is another measure of statistical accuracy, different from either bias or standard error. Let $\hat\theta(\mathbf{y})$ be the 25% trimmed mean and $\mu(F)$ be the mean of F, as in the serum example, and also let $\hat\imath(\mathbf{y})$ be the interquartile range, the distance between the 25th and 75th percentiles of the sample $\mathbf{y} = (x_1, x_2, \cdots, x_n)$. Define

$$R(\mathbf{y}, F) = \frac{\hat\theta(\mathbf{y}) - \mu(F)}{\hat\imath(\mathbf{y})}. \tag{4.6}$$

R is like a Student's t statistic, except that we have substituted the 25% trimmed mean for the sample mean and the interquartile range for the standard deviation.

Suppose we know the 5th and 95th percentiles of $R(\mathbf{y}, F)$, say $\rho^{(.05)}(F)$ and $\rho^{(.95)}(F)$, where the definition of $\rho^{(.05)}(F)$ is

$$\mathrm{Prob}_F\{R(\mathbf{y}, F) < \rho^{(.05)}(F)\} = .05, \tag{4.7}$$

and similarly for $\rho^{(.95)}(F)$. The relationship $\mathrm{Prob}_F\{\rho^{(.05)} \le R < \rho^{(.95)}\} = .90$ combines with definition (4.6) to give a central 90% "t interval" for the mean $\mu(F)$.

$$\mu \in [\hat\theta - \hat\imath\rho^{(.95)}, \hat\theta - \hat\imath\rho^{(.05)}]. \tag{4.8}$$

Of course we do not know $\rho^{(.05)}(F)$ and $\rho^{(.95)}(F)$, but we can approximate them by their bootstrap estimates $\rho^{(.05)}(\hat F)$ and $\rho^{(.95)}(\hat F)$. A bootstrap sample $\mathbf{y}^*$ g[illegible] [illegible] value of (4.6), $R(\mathbf{y}^*, \hat F) = (\hat\theta(\mathbf{y}^*) - \mu(\hat F))/\hat\imath(\mathbf{y}^*)$, where $\hat\imath(\mathbf{y}^*)$ is the interquartile range of the bootstrap data $x_1^*, x_2^*, \cdots, x_n^*$. For any fixed number ρ, the bootstrap estimate of Pro[illegible] based on B bootstrap samples is

$$\#\{R(\mathbf{y}^*(b), \hat F) < \rho\}/B. \tag{4.9}$$

B[illegible] empirical distribution of [illegible] the values of ρ which make (4.9) equal .05 and .95. These approach $\rho^{(.05)}(\hat F)$ and $\rho^{(.95)}(\hat F)$ as $B \to \infty$.

For [illegible] data, $B = 1000$ bootstrap replications gave $\rho^{(.05)}(\hat F) = -.303$ and $\rho^{(.95)}(\hat F) = .078$. Substituting these values into (4.9), and using the observed estimates $\hat\theta = 2.24$, $\hat\imath = 1.40$, gives

$$\mu \in [2.13, 2.66] \tag{4.10}$$

as a central 90% "bootstrap t interval" for the true mean $\mu(F)$. This is considerably shorter than the standard t interval for μ based on 53 degrees of freedom, [illegible] 2.67]. Here $\bar\sigma = .21$ is the usual estimate of standard error (1.3).

Bootstrap confidence intervals are discussed further in Sections 7 and 8. They require more bootstrap replications than do bootstrap standard errors, on the order of $B = 1000$ rather than $B = 50$ or 100. This point is discussed briefly in Section 9.

By now it should be clear that we can use any random variable $R(\mathbf{y}, F)$ to measure accuracy, not just (4.[illegible] [illegible] then estimate $E_F\{R(\mathbf{y}, F)\}$ by its [illegible] $\hat F)\} \doteq \sum_{b=1}^{B} R(\mathbf{y}^*(b), \hat F)/B$. Similarly we can estimate $E_F R(\mathbf{y}, F)^2$ by $E_{\hat F} R(\mathbf{y}^*, \hat F)^2$, etc. Efron (1983) considers the prediction problem, in which [illegible] data is used to construct a pre[illegible] naive estimate of the prediction rule's accuracy is the proportion of correct guesses it makes on its own training set, but this can be greatly ove[illegible] since the prediction rule is explicitly constructed to minimize errors on the training set. In this case, a natural choice of $R(\mathbf{y}, F)$ is the over optimism, the difference between the naive estimate and the actual success rate of the prediction rule for new data. Efron (1983) gives the bootstrap estimate of over optimism and shows that it is closely related to cross-validation, the usual method of estimating over optimism. The paper goes on to show that some modifications of the bootstrap estimate greatly out perform both cross-validation and the bootstrap.

5. MORE COMPLICATED DATA SETS

The bootstrap is not restricted to situations where the data is a simple random sample from a single distribution. Suppose for instance that the data consists of two independent random samples,

$$U_1, U_2, \cdots, U_m \sim F \quad\text{and}\quad V_1, V_2, \cdots, V_n \sim G, \tag{5.1}$$

where F and G are possibly different distributions on the real line. Suppose also that the statistic of interest is the Hodges–Lehmann shift estimate

$$\hat\theta = \text{median}\{V_j - U_i, i = 1, 2, \cdots, m, j = 1, 2, \cdots, n\}. \tag{5.2}$$

Having observed $U_1 = u_1, U_2 = u_2, \cdots, V_n = v_n$, we desire an estimate for $\sigma(F, G)$, the standard error of $\hat\theta$.

The bootstrap estimate of $\sigma(F, G)$ is $\hat\sigma = \sigma(\hat F, \hat G)$, where $\hat F$ is the empirical distribution of $u_1, u_2, \cdots, u_m$, and $\hat G$ is the empirical distribution of $v_1, v_2, \cdots, v_n$. It is easy to modify the Monte Carlo algorithm of Section 2 to numerically evaluate $\hat\sigma$. Let $\mathbf{y} = (u_1, u_2, \cdots, v_n)$ be the observed data vector. A bootstrap sample $\mathbf{y}^* = (u_1^*, u_2^*, \cdots, u_m^*, v_1^*, v_2^*, \cdots, v_n^*)$ consists of a random sample $U_1^*, \cdots, U_m^*$ from $\hat F$ and an independent random sample $V_1^*, \cdots, V_n^*$ from $\hat G$. With only this modification, steps (i) through (iii) of the Monte Carlo algorithm produce $\hat\sigma_B$, (2.4), approaching $\hat\sigma$ as $B \to \infty$.

Table 4 reports on a simulation experiment investigating how well the bootstrap works on this problem. 100 trials of situation (5.1) were run, with $m = 6$, $n = 9$, F and G both Uniform [0, 1]. For each trial, both $B = 100$ and $B = 200$ bootstrap replications were generated. The bootstrap estimate $\hat\sigma_B$ was nearly unbiased for the true standard error $\sigma(F, G) = .167$ for either $B = 100$ or $B = 200$, with a quite small standard deviation from trial to trial. The improvement in going from $B = 100$ to $B = 200$ is too small to show up in this experiment.

In practice, statisticians must often consider quite complicated data structures: time series models, mul-

TABLE 4

Bootstrap estimate of standard error for the Hodges–Lehmann two-sample shift estimate; 100 trials

	Summary statistics for $\hat\sigma_B$		
	Ave	SD	CV
$B = 100$	.165	.030	.18
$B = 200$	.166	.031	.19
True σ	.167		

Note: $m = 6$, $n = 9$; true distributions F and G both uniform [0, 1].

tifactor layouts, sequential sampling, censored and missing data, etc. Fig. 8 illustrates how the bootstrap estimation process proceeds in a general situation. The actual probability mechanism P which generates the observed data $\mathbf{y}$ belongs to some family $\mathbf{P}$ of possible probability mechanism. In the Hodges–Lehmann example, $P = (F, G)$, a pair of distributions on the real line, $\mathbf{P}$ equals the family of all such pairs, and $\mathbf{y} = (u_1, u_2, \cdots, u_m, v_1, v_2, \cdots, v_n)$ is generated by random sampling m times from F and n times from G.

We have a random variable of interest $R(\mathbf{y}, P)$, which depends on both $\mathbf{y}$ and the unknown model P, and we wish to estimate some aspect of the dis- [illegible] Hodges–Lehmann example, $R(\mathbf{y}, P) = \hat\theta(\mathbf{y}) - E_P\{\hat\theta\}$, and we estimated $\sigma(P) = \{E_P R(\mathbf{y}, P)^2\}^{1/2}$, the standard error of $\hat\theta$. As before, the notation E_P indicates expectation when $\mathbf{y}$ is generated according to mechanism P.

We assume that we have some way of estimating the entire probability model P from the data $\mathbf{y}$, producing the estimate called $\hat P$ in Fig. 8. (In the two-sample problem, $\hat P = (\hat F, \hat G)$, the pair of empirical distributions.) *This is the crucial step for the bootstrap.* It can be carried out either parametrically or nonparametrically, by maximum likelihood or by some other estimation technique.

Once we have $\hat P$, we can use Monte Carlo methods to generate bootstrap data sets $\mathbf{y}^*$, according to the same rules by which $\mathbf{y}$ is generated from P. The bootstrap random variable $R(\mathbf{y}^*, \hat P)$ is observable, since we know $\hat P$ as well as $\mathbf{y}^*$, so the distribution of $R(\mathbf{y}^*, \hat P)$ can be found by Monte Carlo sampling. The bootstrap estimate of $E_P R(\mathbf{y}, P)$ is then $E_{\hat P} R(\mathbf{y}^*, \hat P)$, and likewise for estimating any other aspect of $R(\mathbf{y}, P)$'s distribution.

A regression model is a familiar example of a complicated data structure. We observe $\mathbf{y} = (y_1, y_2, \cdots, y_n)$, where

$$y_i = g(\beta, t_i) + \varepsilon_i \quad i = 1, 2, \cdots, n. \tag{5.3}$$

Here β is a vector of unknown parameters we wish to estimate; for each i, t_i is an observed vector of covariates; and g is a known function of β and t_i, for instance $e^{\beta' t_i}$. The ε_i are an iid sample from some unknown distribution F on the real line,

$$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n \sim F, \tag{5.4}$$

where F is usually assumed to be centered at 0 in some sense, perhaps $E\{\varepsilon\} = 0$ or $\text{Prob}\{\varepsilon < 0\} = .5$. The probability model is $P = (\beta, F)$; (5.3) and (5.4) describe the step $P \to \mathbf{y}$ in Fig. 8. The covariates $t_1, t_2, \cdots, t_n$, like the sample size n in the simple problem (1.1), are considered fixed at their observed values.

For every choice of β we have a vector $\mathbf{g}(\beta) = (g(\beta, t_1), g(\beta, t_2), \cdots, g(\beta, t_n))$ of predicted values for $\mathbf{y}$. Having observed $\mathbf{y}$, we estimate β by minimizing some measure of distance between $\mathbf{g}(\beta)$ and $\mathbf{y}$,

$$\hat\beta: \min_{\beta} D(\mathbf{y}, \mathbf{g}(\beta)). \tag{5.5}$$

The most common choice of D is $D(\mathbf{y}, \mathbf{g}) = \sum_{i=1}^n \{y_i - g(\beta, t_i)\}^2$.

How accurate is $\hat\beta$ as an estimate of β? Let $R(\mathbf{y}, P)$ equal the vector $\hat\beta - \beta$. A familiar measure of accuracy is the mean square error matrix

$$\begin{aligned} \Sigma(P) &= E_P(\hat\beta - \beta)(\hat\beta - \beta)' \\ &= E_P R(\mathbf{y}, P) R(\mathbf{y}, P)'. \end{aligned} \tag{5.6}$$

The bootstrap estimate of accuracy $\hat\Sigma = \Sigma(\hat P)$ is obtained by following through Fig. 8.

There is an obvious choice for $\hat P = (\hat\beta, \hat F)$ in this case. The estimate $\hat\beta$ is obtained from (5.5). Then $\hat F$ is the empirical distribution of the residuals,

$$\begin{aligned} \hat F: \text{mass}(1/n) \quad \text{on} \quad \hat\varepsilon_i &= y_i - g(\hat\beta, t_i), \\ i &= 1, \cdots, n. \end{aligned} \tag{5.7}$$

A bootstrap sample $\mathbf{y}^*$ is obtained by following rules (5.3) and (5.4),

$$y_i^* = g(\hat\beta, t_i) + \varepsilon_i^*, \quad i = 1, 2, \cdots, n, \tag{5.8}$$

where $\varepsilon_1^*, \varepsilon_2^*, \cdots, \varepsilon_n^*$ is an iid sample from $\hat F$. Notice that the ε_i^* are independent bootstrap variates, even though the $\hat\varepsilon_i$ are not independent variates in the usual sense.

Each bootstrap sample $\mathbf{y}^*(b)$ gives a bootstrap value $\hat\beta^*(b)$,

$$\hat\beta^*(b): \min_{\beta} D(\mathbf{y}^*(b), \mathbf{g}(\beta)), \tag{5.9}$$

as in (5.5). The estimate

$$\hat\Sigma_B = \frac{\sum_{b=1}^B [\hat\beta^*(b) - \hat\beta^*(\cdot)][\hat\beta^*(b) - \hat\beta^*(\cdot)]'}{B} \tag{5.10}$$

approaches the bootstrap estimate $\hat\Sigma$ as $B \to \infty$. (We could just as well divide by $B - 1$ in (5.10).)

In the case of ordinary least squares regression, where $g(\beta, t_i) = \beta' t_i$ and $D(\mathbf{y}, \mathbf{g}) = \sum_{i=1}^n (y_i - g_i)^2$,

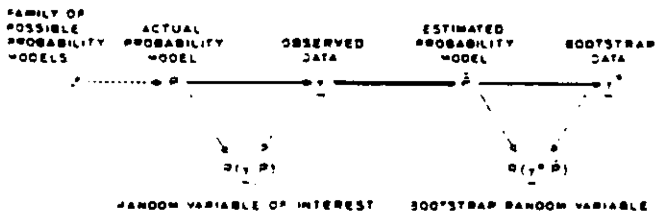


FIG. 8. *A schematic illustration of the bootstrap process for a general probability model P. The expectation of $R(\mathbf{y}, P)$ is estimated by the bootstrap expectation of $R(\mathbf{y}^*, \hat P)$. The double arrow indicates the crucial step in applying the bootstrap.*

Section 7 of Efron (1979a) shows that the bootstrap estimate, $B = \infty$, can be calculated without Monte Carlo sampling, and is

$$(5.11)\qquad \hat{\Sigma} = \hat{\sigma}^2\left(\sum_{i=1}^{n} t_i t_i'\right)^{-1} \quad \left[\hat{\sigma}^2 = \sum_{1}^{n} \frac{\hat{\epsilon}_i^2}{n}\right].$$

This is the usual Gauss–Markov answer, except for the divisor n in the definition of $\hat{\sigma}^2$.

There is another, simpler way to bootstrap a regression problem. We can consider each covariate-response pair $x_i = (t_i, y_i)$ to be a single data point obtained by simple random sampling from a distribution F. If the covariate vector t_i is p-dimensional, F is a distribution on $p + 1$ dimensions. Then we apply the bootstrap as described originally in Section 2 to the data set $x_1, x_2, \cdots, x_n \sim_{\text{iid}} F$.

The two bootstrap methods for the regression problem are asymptotically equivalent, but can perform quite differently in small sample situations. The class of possible probability models P is different for the two methods. The simple method, described last, takes less advantage of the special structure of the regression problem. It does *not* give answer (5.11) in the case of ordinary least squares. On the other hand the simple method gives a trustworthy estimate of $\hat{\beta}$'s variability *even if the regression model* (5.3) *is not correct*. The bootstrap, as outlined in Fig. 5, is very general, but because of this generality there will often be more than one bootstrap solution for a given problem.

As the final example of this section, we discuss *censored data*. The ages of 97 men at a California retirement center, Channing House, were observed either at death (an uncensored observation) or at the time the study ended (a censored observation). The data set $y = \{(x_1, d_1), (x_2, d_2), \cdots, (x_{97}, d_{97})\}$, where x_i was the age of the ith man observed, and

$$d_i = \begin{cases} 1 & \text{if } x_i \text{ uncensored} \\ 0 & \text{if } x_i \text{ censored.} \end{cases}$$

Thus (777, 1) represents a Channing House man observed to die at age 777 months, while (843, 0) represents a man 843 months old when the study ended. His observation could be written as "843+," and in fact d_i is just an indicator for the absence or presence of "+." A full description of the Channing House data appears in Hyde (1980).

A typical data point (X_i, D_i) can be thought of as generated in the following way: a real lifetime X_i^0 is selected randomly according to a survival curve

$$(5.12)\qquad S^0(t) \equiv \operatorname{Prob}\{X_i^0 > t\}, \quad (0 \le t < \infty)$$

and a censoring time W_i is independently selected according to another survival curve

$$(5.13)\qquad R(t) \equiv \operatorname{Prob}\{W_i > t\}, \quad (0 \le t < \infty).$$

The statistician gets to observe

$$(5.14)\qquad X_i = \min\{X_i^0, W_i\}$$

and

$$(5.15)\qquad D_i = \begin{cases} 1 & \text{if } X_i = X_i^0 \\ 0 & \text{if } X_i = W_i. \end{cases}$$

Note: $1 - S^0(t)$ and $1 - R(t)$ are the cumulative distribution functions (cdf) for X_i^0 and W_i, respectively; with censored data it is more convenient to consider survival curves than cdf.

Under assumptions (5.12)–(5.15) there is a simple formula for the nonparametric MLE of $S^0(t)$, called the *Kaplan–Meier estimator* (Kaplan and Meier, 1958). For convenience suppose $x_1 < x_2 < x_3 < \cdots < x_n$, $n = 97$. Then the Kaplan–Meier estimate is

$$(5.16)\qquad \hat{S}^0(t) = \prod_{i=1}^{k_t} \left(\frac{n - i}{n - i + 1}\right)^{d_i},$$

where k_t is the value of k such that $t \in [x_k, x_{k+1})$. In the case of no censoring, $\hat{S}^0(t)$ is equivalent to the observed empirical distribution of $x_1, x_2, \cdots, x_n$, but otherwise (5.16) corrects the empirical distribution to account for censoring. Likewise

$$(5.17)\qquad \hat{R}(t) = \prod_{i=1}^{k_t} \left(\frac{n - i}{n - i + 1}\right)^{1-d_i}$$

is the Kaplan–Meier estimate of the censoring curve $R(t)$.

Fig. 9 shows $\hat{S}^0(t)$ for the Channing House men. It crosses the 50% survival level at $\hat{\theta} = 1044$ months. Call this value the observed median lifetime. We can use the bootstrap to assign a standard error to the observed median.

The probability mechanism is $P = (S^0, R)$; P produces (X_i^0, D_i) according to (5.12)–(5.15), and $y = \{(x_1, d_1), \cdots, (x_n, d_n)\}$ by $n = 97$ independent repetitions of this process. An obvious choice of the estimate $\hat{P}$ in Fig. 8 is $(\hat{S}^0, \hat{R})$, (5.14), (5.15). The rest of

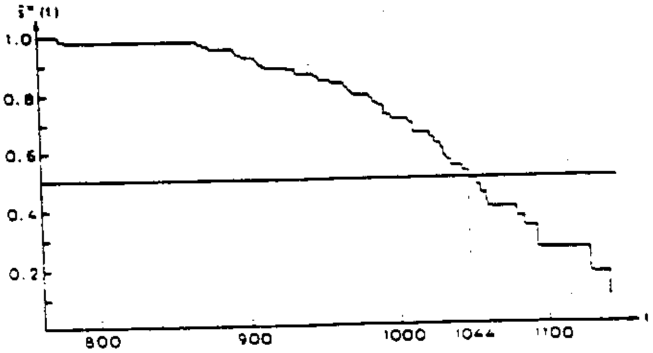


FIG. 9. *Kaplan–Meier estimated survival curve for the Channing House men; t = age in months. The median survival age is estimated to be 1044 months (87 years).*

bootstrap process is automatic: $\hat{S}^0$ and $\hat{R}$ replace S^0 and R in (5.12) and (5.13): n pairs (X_i^*, D_i^*) are independently generated according to rules (5.12)–(5.15), giving the bootstrap data set $\mathbf{y}^* = \{(x_1^*, d_1^*), \cdots, (x_n^*, d_n^*)\}$; and finally the bootstrap Kaplan–Meier curve $\hat{S}^{0*}$ is constructed according to formula (5.16), and the bootstrap observed median $\hat{\theta}^*$ calculated. For the Channing House data, $B = 1600$ bootstrap replications of $\hat{\theta}^*$ gave estimated standard error $\hat{\sigma} = 14.0$ months for $\hat{\theta}$. An estimated bias of 4.1 months was calculated as at (4.4). Efron (1981b) gives a fuller description.

Once again there is a simpler way to apply to bootstrap. Consider each pair $y_i = (x_i, d_i)$ as an observed point obtained by simple random sampling from a bivariate distribution F, and apply the bootstrap as described in Section 2 to the data set $y_1, y_2, \cdots, y_n \sim_{\text{iid}} F$. This method makes no use of the special structure (5.12)–(5.15). Surprisingly, it gives *exactly the same answers* as the more complicated bootstrap method described earlier (Efron, 1981a). This leads to a surprising conclusion: bootstrap estimates of variability for the Kaplan–Meier curve give correct standard errors even when the usual assumptions about the censoring mechanism, (5.12)–(5.15), fail.

6. EXAMPLES WITH MORE COMPLICATED DATA STRUCTURES

Example 1: Autoregressive Time Series Model

This example illustrates an application of the bootstrap to a famous time series.

The data are the Wolfer annual sunspot numbers for the years 1770–1889 (taken from Anderson, 1975). Let the count for the ith year be z_i. After centering the data (replacing z_i by $z_i - \bar{z}$), we fit a first-order autoregressive model

$$z_i = \phi z_{i-1} + \varepsilon_i \tag{6.1}$$

where $\varepsilon_i \sim$ iid $N(0, \sigma^2)$. The estimate $\hat{\phi}$ turned out to be .815 with an estimated standard error, one over the square root of the Fisher information, of .053.

A bootstrap estimate of the standard error of $\hat{\phi}$ can be obtained as follows. Define the residuals $\hat{\varepsilon}_i = z_i - \hat{\phi} z_{i-1}$ for $i = 2, 3, \cdots, 120$. A bootstrap sample $z_1^*, z_2^*, \cdots, z_{120}^*$ is created by sampling $\hat{\varepsilon}_2^*, \hat{\varepsilon}_3^*, \cdots, \hat{\varepsilon}_{120}^*$ with replacement from the residuals, then letting $z_1^* = z_1$, and $z_i^* = \hat{\phi} z_{i-1}^* + \hat{\varepsilon}_i^*$, $i = 2, \cdots, 120$. Finally, after centering the time series $z_1^*, z_2^*, \cdots, z_{120}^*$, $\hat{\phi}^*$ is the estimate of the autoregressive parameter for this new time series. (We could, if we wished, sample the $\hat{\varepsilon}_i^*$ from a fitted normal distribution.)

A histogram of 1000 such bootstrap values $\hat{\phi}_1^*, \hat{\phi}_2^*, \cdots, \hat{\phi}_{1000}^*$ is shown in Fig. 10.

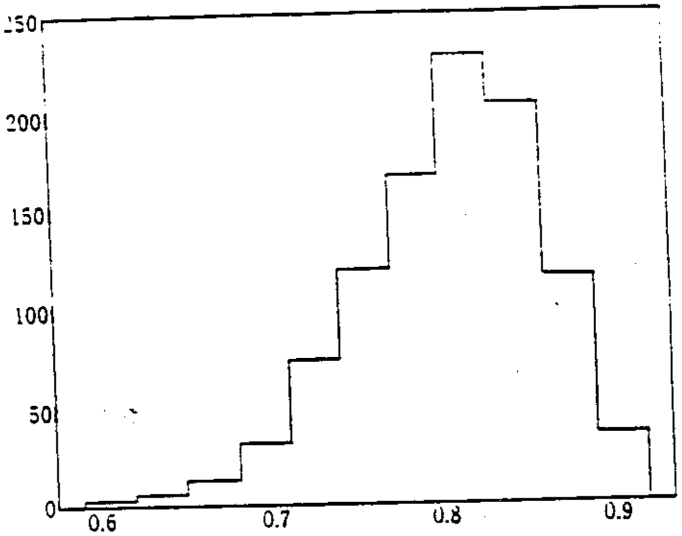


FIG. 10. *Bootstrap histogram of $\hat{\phi}_1^*, \cdots, \hat{\phi}_{1000}^*$ for the Wolfer sunspot data, model (6.1).*

The bootstrap estimate of standard error was .055, agreeing nicely with the usual formula. Note however that the distribution is skewed to the left, so a confidence interval for ϕ might be asymmetric about $\hat{\phi}$ as discussed in Sections 8 and 9.

In bootstrapping the residuals, we have assumed that the first-order autoregressive model is correct. (Recall the discussion of regression models in Section 5.) In fact, the first-order autoregressive model is far from adequate for this data. A fit of second-order autoregressive model

$$z_i = \alpha z_{i-1} + \theta z_{i-2} + \varepsilon_i \tag{6.2}$$

gave estimates $\hat{\alpha} = 1.37$, $\hat{\theta} = -.677$, both with an estimated standard error of .067, based on Fisher information calculations. We applied the bootstrap to this model, producing the histograms for $\alpha_1^*, \cdots, \alpha_{1000}^*$ and $\theta_1^*, \cdots, \theta_{1000}^*$ shown in Figs. 11 and 12, respectively.

The bootstrap standard errors were .070 and .068, respectively, both close to the usual value. Note that the additional term has reduced the skewness of the first coefficient.

Example 2: Estimating a Response Transformation in Regression

Box and Cox (1964) introduced a parametric family for estimating a transformation of the response in a regression. Given regression data $\{(x_1, y_1), \cdots, (x_n, y_n)\}$, their model takes the form

$$z_i(\lambda) = x_i \cdot \beta + \varepsilon_i \tag{6.3}$$

where $z_i(\lambda) = (y_i^\lambda - 1)/\lambda$ for $\lambda \neq 0$ and $\log y_i$ for $\lambda = 0$, and $\varepsilon_i \sim$ iid $N(0, \sigma^2)$. Estimates of λ and β are found by minimizing $\sum_1^n (z_i - x_i \cdot \beta)^2$.

Breiman and Friedman (1985) proposed a nonparametric solution for this problem. Their so called ACE

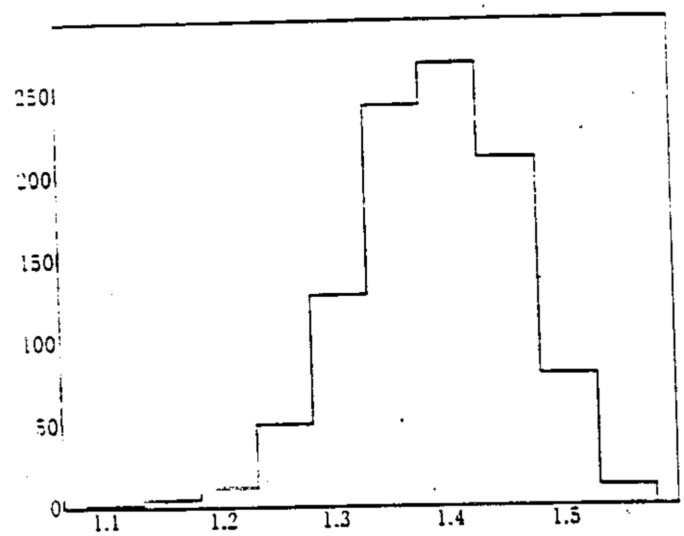


FIG. 11. *Bootstrap histogram of* $\hat{\alpha}^*_1, \cdots, \hat{\alpha}^*_{1000}$ *for the Wolfer sunspot data, model* (6.2).

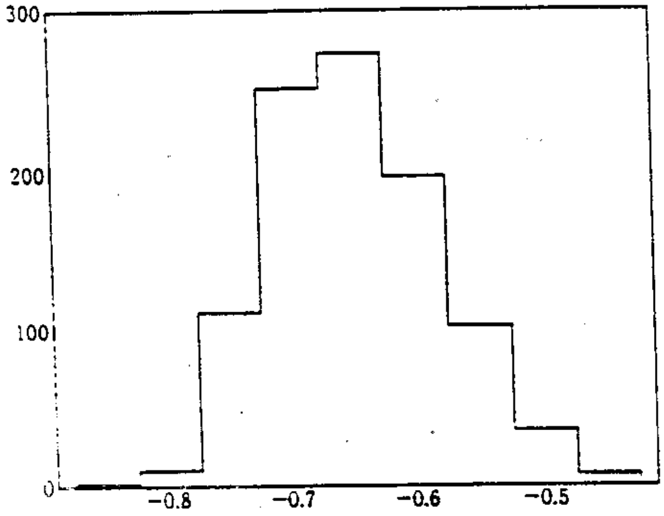


FIG. 12. *Bootstrap histogram of* $\hat{\theta}^*_1, \cdots, \hat{\theta}^*_{1000}$ *for the Wolfer sunspot data, model* (6.2).

(alternating conditional expectation) model generalizes (6.3) to

$$s(y_i) = x_i \cdot \beta + \epsilon_i, \tag{6.4}$$

where $s(\cdot)$ is an unspecified smooth function. (In its most general form, ACE allows for transformations of the covariates as well.) The function $s(\cdot)$ and parameter β are estimated in an alternating fashion, utilizing a nonparametric smoother to estimate $s(\cdot)$.

In the following example, taken from Friedman and Tibshirani (1984), we compare the Box and Cox procedure to ACE and use the bootstrap to assess the variability of ACE.

The data from Box and Cox (1964) consist of a 3 × 3 × 3 experiment on the strength of yarns, the response Y being number of cycles to failure, and the factors length of test specimen (X_1) (250, 300, and 350 mm), amplitude of loading cycle (X_2) (8, 9, or 10 mm), and load (X_3) (40, 45, or 50 g). As in Box and Cox, we treat the factors as quantitive and allow only a linear term for each. Box and Cox found that a logarithmic transformation was appropriate, with their procedure producing a value of −.06 for $\hat{\lambda}$ with an estimated 95% confidence interval of (−.18, .06).

Fig. 13 shows the transformation selected by the ACE algorithm. For comparison, the log function is plotted (normalized) on the same figure.

The similarity is truly remarkable! In order to assess the variability of the ACE curve, we can apply the bootstrap. Since the X matrix in this problem is fixed by design, we resampled from the residuals instead of from the (x_i, y_i) pairs. The bootstrap procedure was the following:

```
Calculate residuals   $\hat{\epsilon}_i = \hat{s}(y_i) - x_i \cdot \hat{\beta}$,   $i = 1, 2, \cdots, n$.
Repeat $B$ times
  Choose a sample   $\hat{\epsilon}^*_1, \cdots, \hat{\epsilon}^*_n$
                    with replacement from   $\hat{\epsilon}_1, \cdots, \hat{\epsilon}_n$
  Calculate   $y^*_i = \hat{s}^{-1}(x_i \cdot \hat{\beta} + \hat{\epsilon}^*_i)$,   $i = 1, 2, \cdots, n$
  Compute   $\hat{s}^*(\cdot)$ = result of ACE algorithm
                    applied to   $(x_1, y^*_1), \cdots, (x_n, y^*_n)$
End
```

The number of bootstrap replications B was 20. Note that the residuals are computed on the $s(\cdot)$ scale, not the y scale, because it is on the $s(\cdot)$ scale that the true residuals are assumed to be approximately iid. The 20 estimated transformations, $\hat{s}^*_1(\cdot), \cdots, \hat{s}^*_{20}(\cdot)$ are shown in Fig. 14.

The tight clustering of the smooths indicates that the original estimate $\hat{s}(\cdot)$ has low variability, especially for smaller values of Y. This agrees qualitatively with

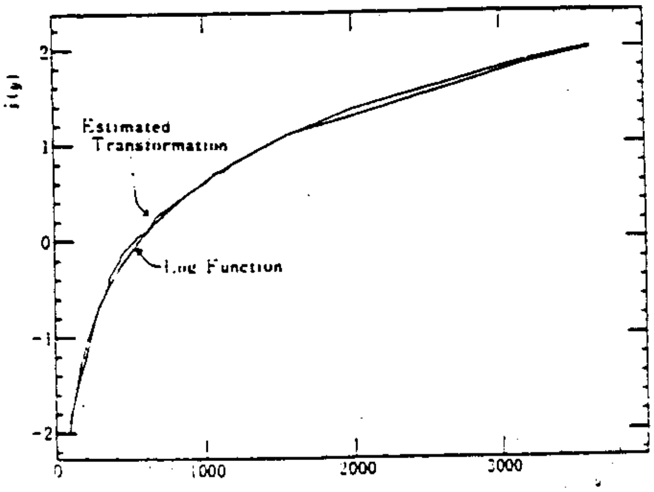


FIG. 13. *Estimated transformation from ACE and the log function for Box and Cox example.*

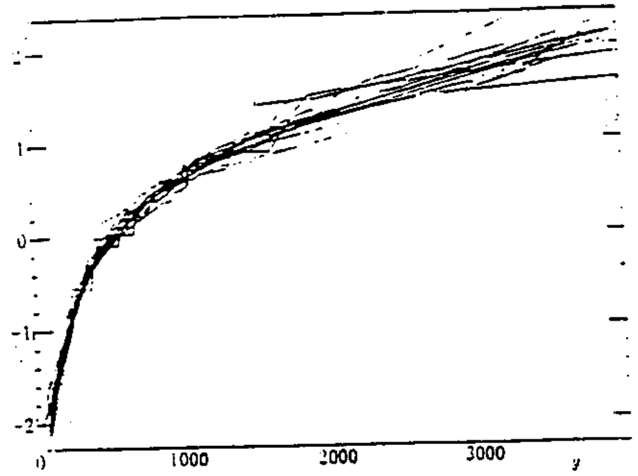


FIG. 14. *Bootstrap replications of ACE transformations for Box and Cox example.*

the short confidence interval for λ in the Box and Cox analysis.

7. BOOTSTRAP CONFIDENCE INTERVALS

This section presents three closely related methods of using the bootstrap to set confidence intervals. The discussion is in terms of simple parametric models, where the logical basis of the bootstrap methods is easiest to see. Section 8 extends the methods to multiparameter and nonparametric models.

We have discussed obtaining $\hat\sigma$, the estimated standard error of an estimator $\hat\theta$. In practice, $\hat\theta$ and $\hat\sigma$ are usually used together to form the approximate confidence interval $\theta \in \hat\theta \pm \hat\sigma z^{(\alpha)}$, (1.7), where $z^{(\alpha)}$ is the $100 \cdot \alpha$ percentile point of a standard normal distribution. The interval (1.7) is claimed to have approximate coverage probability $1 - 2\alpha$. For the law school example of Section 2, the values $\hat\theta = .776$, $\hat\sigma = .115$, $z^{(.05)} = -1.645$, give $\theta \in [.587, .965]$ as an approximate 90% central interval for the true correlation coefficient.

We will call (1.7) the *standard interval for θ*. When working within parametric families like the bivariate normal, $\hat\sigma$ in (1.7) is usually obtained by differentiating the log likelihood function, see Section 5a of Rao (1973), although in the context of this paper we might prefer to use the parametric bootstrap estimate of σ, e.g., $\hat\sigma_{\text{NORM}}$ in Section 2.

The standard intervals are an immensely useful statistical tool. They have the great virtue of being automatic: a computer program can be written which produces (1.7) directly from the data y and the form of the density function for y, with no further input required from the statistician. Nevertheless the standard intervals can be quite inaccurate as Table 5 shows. The standard interval (1.7), using $\hat\sigma_{\text{NORM}}$, (2.5), is strikingly different from the exact normal theory interval based on the assumption of a bivariate normal sampling distribution F.

TABLE 5

Exact and approximate central 90% confidence intervals for θ, the true correlation coefficient, from the law school data of Fig. 1

1. Exact (normal theory)	[.496, .898]	R/L = .44
2. Standard (1.7)	[.587, .965]	R/L = 1.00
3. Transformed standard	[.508, .907]	R/L = .49
4. Parametric bootstrap (BC)	[.488, .900]	R/L = .43
5. Nonparametric bootstrap (BC$_a$)	[.43, .92]	R/L = .42

Note: R/L = ratio of right side of interval, measured from $\hat\theta = .776$, to left side. The exact interval is strikingly asymmetric about $\hat\theta$. Section 8 discusses the nonparametric method of line 5.

In this case, it is well known that it is better to make the transformation $\hat\phi = \tanh^{-1}(\hat\theta)$, $\phi = \tanh^{-1}(\theta)$, apply (1.7) on the ϕ scale, and then transform back to the θ scale. The resulting interval, line 3 of Table 5, is moved closer to the exact interval. However, there is nothing automatic about the $\tanh^{-1}$ transformation. For a different statistic from the correlation coefficient or a different distributional family from the bivariate normal, we might very well need other tricks to make (1.7) perform satisfactorily.

The bootstrap can be used to produce approximate confidence intervals in an automatic way. The following discussion is abridged from Efron (1984 and 1985) and Efron (1982a, Chapter 10). Line 4 of Table 5 shows that the parametric bootstrap interval for the correlation coefficient θ is nearly identical with the exact interval. "Parametric" in this case means that the bootstrap algorithm begins from the bivariate normal MLE $\hat F_{\text{NORM}}$, as for the normal theory curve of Fig. 2. This good performance is no accident. The bootstrap method used in line 4 in effect transforms $\hat\theta$ to the best (most normal) scale, finds the appropriate interval, and transforms this interval back to the θ scale. All of this is done automatically by the bootstrap algorithm, without requiring special intervention from the statistician. The price paid is a large amount of computing, perhaps $B = 1000$ bootstrap replications, as discussed in Section 10.

Define $\hat G(s)$ to be the parametric bootstrap cdf of $\hat\theta^*$,

$$\hat G(s) = \text{Prob}_*\{\hat\theta^* < s\}, \tag{7.1}$$

where Prob$_*$ indicates probability computed according to the bootstrap distribution of $\hat\theta^*$. In Fig. 2 $\hat G(s)$ is obtained by integrating the normal theory curve. We will present three different kinds of bootstrap confidence intervals in order of increasing generality. All three methods use percentiles of $\hat G$ to define the confidence interval. They differ in which percentiles are used.

The simplest method is to take $\theta \in [\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1-\alpha)]$ as an approximate $1 - 2\alpha$ central interval for θ. This is called the *percentile method* in Section 10.4 of Efron (1982a). The percentile method interval is just the interval between the $100 \cdot \alpha$ and $100 \cdot (1-\alpha)$ percentiles of the bootstrap distribution of $\hat{\theta}^*$.

We will use the notation $\theta[\alpha]$ for the α level endpoint of an approximate confidence interval for θ, so $\theta \in [\theta[\alpha], \theta[1-\alpha]]$ is the central $1 - 2\alpha$ interval. Subscripts will be used to indicate the various different methods. The percentile interval has endpoints

$$\theta_P[\alpha] \equiv \hat{G}^{-1}(\alpha). \tag{7.2}$$

This compares with the standard interval,

$$\theta_S[\alpha] = \hat{\theta} + \hat{\sigma} z^{(\alpha)}. \tag{7.3}$$

Lines 1 and 2 of Table 6 summarize these definitions.

Suppose the bootstrap cdf $\hat{G}$ is perfectly normal, say

$$\hat{G}(s) = \Phi((s - \hat{\theta})/\hat{\sigma}), \tag{7.4}$$

where $\Phi(s) = \int_{-\infty}^{s} (2\pi)^{-1/2} e^{-t^2/2}\, dt$, the standard normal cdf. In other words, suppose that $\hat{\theta}^*$ has bootstrap distribution $N(\hat{\theta}, \hat{\sigma}^2)$. In this case the standard method and the percentile method agree, $\theta_S[\alpha] = \theta_P[\alpha]$. In situations like that of Fig. 2, where $\hat{G}$ is markedly nonnormal, the standard interval is quite different from (7.2). Which is better?

To answer this question, consider the simplest possible situation, where for all θ

$$\hat{\theta} \sim N(\theta, \sigma^2). \tag{7.5}$$

That is, we have a single unknown parameter θ with no nuisance parameters, and a single summary statistic $\hat{\theta}$ normally distributed about θ with constant standard error σ. In this case the parametric bootstrap cdf is given by (7.4), so $\theta_S[\alpha] = \theta_P[\alpha]$. (The bootstrap estimate $\hat{\sigma}$ equals σ.)

Suppose though that instead of (7.5) we have, for all θ,

$$\hat{\phi} \sim N(\phi, \tau^2), \tag{7.6}$$

for some monotone transformation $\hat{\phi} = g(\hat{\theta})$, $\phi = g(\theta)$, where τ is a constant. In the correlation coefficient example the function g was $\tanh^{-1}$. The standard limits (7.2) can now be grossly inaccurate. However it is easy to verify that the percentile limits (7.2) are still correct. "Correct" here means that (7.2) is the mapping of the obvious interval for ϕ, $\hat{\phi} \pm \tau z^{(\alpha)}$, back to the θ scale, $\theta_P[\alpha] = g^{-1}(\hat{\phi} + \tau z^{(\alpha)})$. It is also correct in the sense of having exactly the claimed converge probability $1 - 2\alpha$.

Another way to state things is that the percentile intervals are transformation invariant,

$$\phi_P[\alpha] = g(\theta_P[\alpha]) \tag{7.7}$$

for any monotone transformation g. This implies that if the percentile intervals are correct on some transformed scale $\phi = g(\theta)$, then they must also be correct on the original scale θ. The statistician does not need to know the normalizing transformation g, only that it exists. Definition (7.2) automatically takes care of the bookkeeping involved in the use of normalizing transformations for confidence intervals.

Fisher's theory of maximum likelihood estimation says that we are always in situation (7.5) to a first order of asymptotic approximation. However, we are also in situation (7.6), for any choice of g, to the same order of approximation. Efron (1984 and 1985) uses higher order asymptotic theory to differentiate between the standard and bootstrap intervals. It is the higher order asymptotic terms which often make exact intervals strongly asymmetric about the MLE $\hat{\theta}$ as in Table 5. The bootstrap intervals are effective at capturing this asymmetry.

The percentile method automatically incorporates normalizing transformations, as in going from (7.5)–(7.6). It turns out that there are two other important ways that assumption (7.5) can be misleading, the first of which relates to possible bias in $\hat{\theta}$. For example consider $f_\theta(\hat{\theta})$, the family of densities for the observed correlation coefficient $\hat{\theta}$ when sampling $n = 15$ times from a bivariate normal distribution with true corre-

TABLE 6

Four methods of setting approximate confidence intervals for a real valued parameter θ

Method	Abbreviation	α level endpoint	Correct if	
1. Standard	$\theta_S[\alpha]$	$\hat{\theta} + \hat{\sigma} z^{(\alpha)}$	$\hat{\theta} \sim N(\theta, \sigma^2)$	σ constant
			There exists monotone transformation $\hat{\phi} = g(\hat{\theta})$, $\phi = g(\theta)$ such that:	
2. Percentile	$\theta_P[\alpha]$	$\hat{G}^{-1}(\alpha)$	$\hat{\phi} \sim N(\phi, \tau^2)$	τ constant
3. Bias-corrected	$\theta_{BC}[\alpha]$	$\hat{G}^{-1}(\Phi\{2z_0 + z^{(\alpha)}\})$	$\hat{\phi} \sim N(\phi - z_0\tau, \tau^2)$	z_0, τ constant
4. BC$_a$	$\theta_{BC_a}[\alpha]$	$\hat{G}^{-1}\left(\Phi\left\{z_0 + \frac{(z_0 + z^{(\alpha)})}{1 - a(z_0 + z^{(\alpha)})}\right\}\right)$	$\hat{\phi} \sim N(\phi - z_0\tau_\phi, \tau_\phi^2)$ where $\tau_\phi = 1 + a\phi$	z_0, a constant

Note: Each method is correct under more general assumptions than its predecessor. Methods 2, 3, and 4 are defined in terms of the percentiles of $\hat{G}$, the bootstrap distribution (7.1).

lation θ. In fact it is easy to see that no monotone mapping $\hat{\phi} = g(\hat{\theta})$, $\phi = g(\theta)$ transforms this family to $\hat{\phi} \sim N(\phi, \tau^2)$, as in (7.6). If there were such a g, then $\text{Prob}_\theta\{\hat{\theta} < \theta\} = \text{Prob}_\phi\{\hat{\phi} < \phi\} = .50$, but for $\theta = .776$ integrating the density function $f_{.776}(\hat{\theta})$ gives $\text{Prob}_{\theta=.776}\{\hat{\theta} < \theta\} = .431$.

The *bias-corrected percentile method* (BC method), line 3 of Table 6, makes an adjustment for this type of bias. Let

$$z_0 \equiv \Phi^{-1}\{\hat{G}(\hat{\theta})\}, \tag{7.8}$$

where Φ^{-1} is the inverse function of the standard normal cdf. The BC method has α level endpoint

$$\theta_{\text{BC}}[\alpha] \equiv \hat{G}^{-1}(\Phi\{2z_0 + z^{(\alpha)}\}). \tag{7.9}$$

Note: if $\hat{G}(\hat{\theta}) = .50$, that is if half of the bootstrap distribution of $\hat{\theta}^*$ is less than the observed value $\hat{\theta}$, then $z_0 = 0$ and $\theta_{\text{BC}}[\alpha] = \theta_P[\alpha]$. Otherwise definition (7.9) makes a bias correction.

Section 10.7 of Efron (1982a) shows that the BC interval for θ is exactly correct if

$$\hat{\phi} \sim N(\phi - z_0\tau, \tau^2) \tag{7.10}$$

for some monotone transformation $\hat{\phi} = g(\hat{\theta})$, $\phi = g(\theta)$ and some constant z_0. It does not look like (7.10) is much more general than (7.6), but in fact the bias correction is often important.

In the example of Table 5, the percentile method (7.2) gives central 90% interval [.536, .911] compared to the BC interval [.488, .900] and the exact interval [.496, .898]. By definition the endpoints of the exact interval satisfy

$$\begin{aligned} \text{Prob}_{\theta=.496}\{\hat{\theta} > .776\} &= .05 \\ &= \text{Prob}_{\theta=.898}\{\hat{\theta} < .776\}. \end{aligned} \tag{7.11}$$

The corresponding quantities for the BC endpoints are

$$\begin{aligned} \text{Prob}_{\theta=.488}\{\hat{\theta} > .776\} &= .0465, \\ \text{Prob}_{\theta=.900}\{\hat{\theta} < .776\} &= .0475, \end{aligned} \tag{7.12}$$

compared to

$$\begin{aligned} \text{Prob}_{\theta=.536}\{\hat{\theta} > .776\} &= .0725, \\ \text{Prob}_{\theta=.911}\{\hat{\theta} < .776\} &= .0293. \end{aligned} \tag{7.13}$$

for the percentile endpoints. The bias correction is quite important in equalizing the error probabilities at the two endpoints. If z_0 can be approximated accurately (as mentioned in Section 9), then it is preferable to use the BC intervals.

Table 7 shows a simple example where the BC method is less successful. The data consists of the single observation $\hat{\theta} \sim \theta(\chi^2_{19}/19)$, the notation indicating an unknown scale parameter θ times a random variable with distribution $\chi^2_{19}/19$. (This definition makes $\hat{\theta}$ unbiased for θ.) A confidence interval is desired for the scale parameter θ. In this case the BC interval based on $\hat{\theta}$ is a definite improvement over the standard interval (1.7), but goes only about half as far as it should toward achieving the asymmetry of the exact interval.

TABLE 7

Central 90% confidence intervals for θ having observed $\hat{\theta} \sim \theta(\chi^2_{19}/19)$

1. Exact	$[.631 \cdot \hat{\theta}, 1.88 \cdot \hat{\theta}]$	R/L = 2.38
2. Standard (1.7)	$[.466 \cdot \hat{\theta}, 1.53 \cdot \hat{\theta}]$	R/L = 1.00
3. BC (7.9)	$[.580 \cdot \hat{\theta}, 1.69 \cdot \hat{\theta}]$	R/L = 1.64
4. BC$_a$ (7.15)	$[.630 \cdot \hat{\theta}, 1.88 \cdot \hat{\theta}]$	R/L = 2.37
5. Nonparametric BC$_a$	$[.640 \cdot \hat{\theta}, 1.68 \cdot \hat{\theta}]$	R/L = 1.88

Note: The exact interval is sharply skewed to the right of $\hat{\theta}$. The BC method is only a partial improvement over the standard interval. The BC$_a$ interval, $a = .108$, agrees almost perfectly with the exact interval.

It turns out that the parametric family $\hat{\theta} \sim \theta(\chi^2_{19}/19)$ cannot be transformed into (7.10), not even approximately. The results of Efron (1982b) show that there does exist a monotone transformation g such that $\hat{\phi} = g(\hat{\theta})$, $\phi = g(\theta)$ satisfy to a high degree of approximation

$$\hat{\phi} \sim N(\phi - z_0\tau_\phi, \tau_\phi^2) \quad (\tau_\phi = 1 + a\phi). \tag{7.14}$$

The constants in (7.14) are $z_0 = .1082$, $a = .1077$.

The BC$_a$ method (Efron, 1984), line 4 of Table 6, is a method of assigning bootstrap confidence intervals which are exactly right for problems which can be mapped into form (7.14). This method has α level endpoint

$$\theta_{\text{BC}_a}[\alpha] \equiv \hat{G}^{-1}\left(\Phi\left\{z_0 + \frac{z_0 + z^{(\alpha)}}{1 - a(z_0 + z^{(\alpha)})}\right\}\right). \tag{7.15}$$

If $a = 0$ then $\theta_{\text{BC}_a}[\alpha] = \theta_{\text{BC}}[\alpha]$, but otherwise the BC$_a$ intervals can be a substantial improvement over the BC method as shown in Table 7.

The constant z_0 in (7.15) is given by $z_0 = \Phi^{-1}\{\hat{G}(\hat{\theta})\}$, (7.8), and so can be computed directly from the bootstrap distribution. How do we know a? It turns out that in one-parameter families $f_\theta(\hat{\theta})$, a good approximation is

$$a \doteq \frac{\text{SKEW}_{\theta=\hat{\theta}}(\dot{L}_\theta(t))}{6}, \tag{7.16}$$

where $\text{SKEW}_{\theta=\hat{\theta}}(\dot{L}_\theta(t))$ is the skewness at parameter value $\theta = \hat{\theta}$ of the score statistic $\dot{L}_\theta(t) = (\partial/\partial\theta)\log f_\theta(t)$. For $\hat{\theta} \sim \theta(\chi^2_{19}/19)$ this gives $a \doteq .1081$, compared to the actual value $a = .1077$ derived in Efron (1984). For the normal theory correlation family of Table 5 $a \doteq 0$ which explains why the BC method, which takes $a = 0$, words so well there.

The advantage of formula (7.18) is that we need not know the transformation g leading to (7.14) in order to approximate a. In fact $\theta_{BC_a}[\alpha]$, like $\theta_{BC}[\alpha]$ and $\theta_P[\alpha]$, is transformation invariant, as in (7.7). Like the bootstrap methods, the BC_a intervals are computed directly from the form of the density function $f_\theta(\cdot)$, for θ near $\hat{\theta}$.

Formula (7.16) applies to the case where θ is the only parameter. Section 8 briefly discusses the more challenging problem of setting confidence intervals for a parameter θ in a multiparameter family, and also in nonparametric situations where the number of nuisance parameters is effectively infinite.

To summarize this section, the progression from the standard intervals to the BC_a method is based on a series of increasingly less restrictive assumptions, as shown in Table 6. Each successive method in Table 6 requires the statistician to do a greater amount of computation; first the bootstrap distribution $\hat{G}$, then the bias correction constant z_0, and finally the constant a. However, all of these computations are algorithmic in character, and can be carried out in an automatic fashion.

Chapter 10 of Efron (1982a) discusses several other ways of using the bootstrap to construct approximate confidence intervals, which will not be presented here. One of these methods, the "bootstrap t," was used in the blood serum example of Section 4.

8. NONPARAMETRIC AND MULTIPARAMETER CONFIDENCE INTERVALS

Section 7 focused on the simple case $\hat{\theta} \sim f_\theta$, where we have only a real valued parameter θ and a real valued summary statistic $\hat{\theta}$ from which we are trying to construct a confidence interval for θ. Various favorable properties of the bootstrap confidence intervals were demonstrated in the simple case, but of course the simple case is where we least need a general method like the bootstrap.

Now we will discuss the more common situation where there are nuisance parameters besides the parameter of interest θ; or even more generally the nonparametric case, where the number of nuisance parameters is effectively infinite. The discussion is limited to a few brief examples. Efron (1984 and 1985) develops the theoretical basis of bootstrap approximate confidence intervals for complicated situations, and gives many more examples. The word "approximate" is important here since exact nonparametric confidence intervals do not exist for most parameters (see Bahadur and Savage, 1956).

Example 1. Ratio Estimation

TABLE 8

Central 90% confidence intervals for $\theta = \eta_2/\eta_1$ and for $\phi = 1/\theta$ having observed $(y_1, y_2) = (8, 4)$ from a bivariate normal distribution $\mathbf{y} \sim N_2(\eta, \mathrm{I})$

	For θ	For ϕ
1. Exact (Fieller)	[.29, .76]	[1.32, 3.50]
2. Parametric boot (BC)	[.29, .76]	[1.32, 3.50]
3. Standard (1.7)	[.27, .73]	[1.08, 2.92]
MLE	$\hat{\theta} = .5$	$\hat{\phi} = 2$

Note: The BC intervals, line 2, are based on the parametric bootstrap distribution of $\hat{\theta} = y_2/y_1$.

The data consists of $\mathbf{y} = (y_1, y_2)$, assumed to come from a bivariate normal distribution with unknown mean vector η and covariance matrix the identity,

$$\mathbf{y} \sim N_2(\eta, \mathrm{I}). \tag{8.1}$$

The parameter of interest, for which we desire a confidence interval, is the ratio

$$\theta = \eta_2/\eta_1. \tag{8.2}$$

Fieller (1954) provided well known exact intervals for θ in this case. The Fieller intervals are based on a clever trick, which seems very special to situation (8.1), (8.2).

Table 8 shows Fieller's central 90% interval for θ having observed $\mathbf{y} = (8, 4)$. Also shown is the Fieller interval for $\phi = 1/\theta = \eta_1/\eta_2$, which equals $[.76^{-1}, .29^{-1}]$, the obvious transformation of the interval for θ. The standard interval (1.7) is satisfactory for θ, but not for ϕ. Notice that the standard interval does not transform correctly from θ to ϕ.

Line 2 shows the BC intervals based on applying definitions (7.8) and (7.9) to the parametric bootstrap distribution of $\hat{\theta} = y_2/y_1$ (or $\hat{\phi} = y_1/y_2$). This is the distribution of $\hat{\theta}^* = y_2^*/y_1^*$ when sampling $\mathbf{y}^* = (y_1^*, y_2^*)$ from $\hat{F}_{\mathrm{NORM}} \sim N_2((y_1, y_2), \mathrm{I})$. The bootstrap intervals transform correctly, and in this case they agree with the exact interval to three decimal places.

Example 2. Product of Normal Means

For most multiparameter situations, there do not exist exact confidence intervals for a single parameter of interest. Suppose for instance that (8.2) is changed to

$$\theta = \eta_1\eta_2, \tag{8.3}$$

still assuming (8.1). Table 9 shows approximate intervals for θ, and also for $\phi = \theta^2$, having observed $\mathbf{y} = (2, 4)$. The "almost exact" intervals are based on an analog of Fieller's argument (Efron, 1985), which with suitable care can be carried through to a high degree of accuracy. Once again, the parametric BC intervals are a close match to line 1. The fact that the standard intervals do not transform correctly is particularly obvious here.

TABLE 9

Central 90% confidence intervals for $\theta = \eta_1\eta_2$ and $\phi = \theta^2$ having observed $\mathbf{y} = (2, 4)$, where $\mathbf{y} \sim N_2(\eta, \mathrm{I})$

	For θ	For ϕ
1. Almost exact	[1.77, 17.03]	[3.1, 290.0]
2. Parametric boot (BC)	[1.77, 17.12]	[3.1, 239.1]
3. Standard (1.7)	[0.64, 15.36]	[−53.7, 181.7]
MLE	$\hat{\theta} \doteq 8$	$\hat{\phi} = 64$

Note: The almost exact intervals are based on the high order approximation theory of Efron (1985). The BC intervals of line 2 are based on the parametric bootstrap distribution of $\hat{\theta} = y_1 y_2$.

The good performance of the parametric BC intervals is not accidental. The theory developed in Efron (1985) shows that the BC intervals, based on bootstrapping the MLE $\hat{\theta}$, agree to high order with the almost exact intervals in the following class of problems: the data $\mathbf{y}$ comes from a multiparameter family of densities $f_\eta(\mathbf{y})$, both $\mathbf{y}$ and η k-dimensional vectors; the real valued parameter of interest θ is a smooth function of η, $\theta = t(\eta)$; and the family $f_\eta(\mathbf{y})$ can be transformed to multivariate normality, say

$$g(\mathbf{y}) \sim N_k(h(\eta), \mathrm{I}), \tag{8.4}$$

by some one-to-one transformations g and h.

Just as in Section 7, it is not necessary for the statistician to know the normalizing transformations g and h, only that they exist. The BC intervals are obtained directly from the original densities f_η: we find $\hat{\eta} = \hat{\eta}(\mathbf{y})$, the MLE of η; sample $\mathbf{y}^* \sim f_{\hat{\eta}}$; compute $\hat{\theta}^*$, the bootstrap MLE of θ; calculate $\hat{G}$, the bootstrap cdf of $\hat{\theta}^*$, usually by Monte Carlo sampling, and finally apply definitions (7.8) and (7.9). This process gives the same interval for θ whether or not the transformation to form (8.4) has been made.

Not all problems can be transformed as in (8.4) to a normal distribution with constant covariance. The case considered in Table 7 is a one-dimensional counter example. As a result the BC intervals do not always work as well as in Tables 8 and 9, although they usually improve on the standard method. However, in order to take advantage of the BC$_a$ method, which is based on more general assumptions, we need to be able to calculate the constant a.

Efron (1984) gives expressions for "a" generalizing (7.16) to multiparameter families, and also to nonparametric situations. If (8.4) holds, then "a" will have value zero, and the BC$_a$ method reduces to the BC case. Otherwise the two intervals differ.

Here we will discuss only the nonparametric situation: the observed data $\mathbf{y} = (x_1, x_2, \cdots, x_n)$ consists of iid observations $X_1, X_2, \cdots, X_n \sim F$, where F can be *any* distribution on the sample space $\mathcal{X}$; we want a confidence interval for $\theta = t(F)$, some real valued functional of F; and the bootstrap interval are based on bootstrapping $\hat{\theta} = t(\hat{F})$, which is the nonparametric MLE of θ. In this case a good approximation to the constant a is given in terms of the empirical influence function U_i^0, defined in Section 10 at (10.11),

$$a \doteq \frac{1}{6} \frac{\sum_{i=1}^{n} (U_i^0)^3}{\{\sum_{i=1}^{n} (U_i^0)^2\}^{3/2}}. \tag{8.5}$$

This is a convenient formula, since it is easy to numerically evaluate the U_i^0 by simply substituting a small value of θ into (10.11).

Example 3. The Law School Data

For $\hat{\theta}$ the correlation coefficient, the values of U_i^0 corresponding to the 15 data points shown in Fig. 1 are −1.507, .168, .273, .004, .525, −.049, −.100, .477, .310, .004, −.526, −.091, .434, .125, −.048. (Notice how influential law school 1 is.) Formula (8.5) gives $a \doteq -.0817$. $B = 100{,}000$ bootstrap replications, about 100 times more than was actually necessary (see Section 10), gave $z_0 = -.0927$, and the central 90% interval $\theta \in [.43, .92]$ shown in Table 5. The nonparametric BC$_a$ interval is quite reasonable in this example, particularly considering that there is no guarantee that the true law school distribution F is anywhere near bivariate normal.

Example 4. Mouse Leukemia Data (the First Example in Section 3)

The standard central 90% interval for β in formula (3.1) is [.835, 2.18]. The bias correction constant $z_0 \doteq .0275$, giving BC interval [1.00, 2.39]. This is shifted far right of the standard interval, reflecting the long right tail of the bootstrap histogram seen in Fig. 3. We can calculate "a" from (8.5), considering each of the $n = 42$ data points to be a triple (y_i, x_i, δ_i): $a \doteq -.152$. Because a is negative, the BC$_a$ interval is shifted back to the left, equaling [.788, 2.10]. This contrasts with the law school example, where a, z_0, and the skewness of the bootstrap distribution added to each other rather than cancelling out, resulting in a BC$_a$ interval much different from the standard interval.

Efron (1984) provides some theoretical support for the nonparametric BC$_a$ method. However the problem of setting approximate nonparametric confidence intervals is still far from well understood, and all methods should be interpreted with some caution. We end this section with a cautionary example.

Example 5. The Variance

Suppose $\mathcal{X}$ is the real line, and $\theta = \mathrm{Var}_F X$, the variance. Line 5 of Table 2 shows the result of applying the nonparametric BC$_a$ method to data sets $x_1, x_2, \cdots, x_{20}$ which were actually iid samples from a $N(0, 1)$ distribution. The number .640 for example is the average of $\hat{\theta}_{\mathrm{BC}}[.05]/\hat{\theta}$ over 40 such data sets. $B = 4000$ bootstrap replications per data set. The upper limit $1.68 \cdot \hat{\theta}$ is noticeably small, as pointed out by Schenker (1985). The reason is simple: the nonparametric bootstrap distribution of $\hat{\theta}^*$ has a short upper tail, compared to the parametric bootstrap distribution which is a scaled χ^2_{19} random variable. The results of Beran (1984), Bickel and Freedman (1981), and Singh (1981) show that the nonparametric bootstrap distribution is highly accurate asymptotically, but of course that is not a guarantee of good small sample behavior. Bootstrapping from a smoothed version of $\hat{F}$, as in lines 3, 4, and 5 of Table 2 alleviates the problem in this particular example.

9. BOOTSTRAP SAMPLE SIZES

How many bootstrap replications must we take? Consider the standard error estimate $\hat{\sigma}_B$ based on B bootstrap replications, (2.4). As $B \to \infty$, $\hat{\sigma}_B$ approaches $\hat{\sigma}$, the bootstrap estimate of standard error as originally defined in (2.3). Because $\hat{F}$ does not estimate F perfectly, $\hat{\sigma} = \sigma(\hat{F})$ will have a non-zero coefficient of variation for estimating the true standard error $\sigma = \sigma(F)$; $\hat{\sigma}_B$ will have a larger CV because of the randomness added by the Monte Carlo bootstrap sampling.

It is easy to derive the following approximation,

$$\text{(9.1)} \qquad \mathrm{CV}(\hat{\sigma}_B) \doteq \left\{ \mathrm{CV}(\hat{\sigma})^2 + \frac{E\{\hat{\delta}\} + 2}{4B} \right\}^{1/2},$$

where $\hat{\delta}$ is the kurtosis of the bootstrap distribution of $\hat{\theta}^*$, given the data y, and $E\{\hat{\delta}\}$ its expected value averaged over y. For typical situations, $\mathrm{CV}(\hat{\sigma})$ lies between .10 and .30. For example, if $\hat{\theta} = \bar{x}$, $n = 20$, $x_i \sim_{\text{iid}} N(0, 1)$, then $\mathrm{CV}(\hat{\sigma}) \doteq .16$.

Table 10 shows $\mathrm{CV}(\hat{\sigma}_B)$ for various values of B and $\mathrm{CV}(\hat{\sigma})$, assuming $E\{\hat{\delta}\} = 0$ in (9.1). For values of $\mathrm{CV}(\hat{\sigma}) > .10$, *there is little improvement past* $B = 100$. In fact B as small as 25 gives reasonable results. Even smaller values of B can be quite informative, as we saw in the Stanford Heart Transplant Data (Fig. 7 of Section 3).

TABLE 10

Coefficient of variation of $\hat{\sigma}_B$, the bootstrap estimate of standard error based on B Monte Carlo replications, as a function of B and $\mathrm{CV}(\hat{\sigma})$, *the limiting CV as* $B \to \infty$

		$B \to$				
		25	50	100	200	∞
$\mathrm{CV}(\hat{\sigma})$	.25	.29	.27	.26	.25	.25
↓	.20	.24	.22	.21	.21	.20
	.15	.21	.18	.17	.16	.15
	.10	.17	.14	.12	.11	.10
	.05	.15	.11	.09	.07	.05
	0	.14	.10	.07	.05	0

Note: Based on (9.1), assuming $E\{\hat{\delta}\} = 0$.

The situation is quite different for setting bootstrap confidence intervals. The calculations of Efron (1984), Section 8, show that $B = 1000$ is a rough minimum for the number of Monte Carlo bootstraps necessary to compute the BC or BC$_a$ intervals. Somewhat smaller values, say $B = 250$, can give a useful percentile interval, the difference being that then the constant z_0 need not be computed. Confidence intervals are a fundamentally more ambitious measure of statistical accuracy than standard errors, so it is not surprising that they require more computational effort.

10. THE JACKKNIFE AND THE DELTA METHOD

This section returns to the simple case of assigning a standard error to $\hat{\theta}(y)$, where $\mathbf{y} = (x_1, \cdots, x_n)$ is obtained by random sampling from a single unknown distribution, $X_1, \cdots, X_n \sim_{\text{iid}} F$. We will give another description of the bootstrap estimate $\hat{\sigma}$, which illustrates the bootstrap's relationship to older techniques of assigning standard errors, like the jackknife and the delta method.

For a given bootstrap sample $\mathbf{y}^* = (x_1^*, \cdots, x_n^*)$, as described in step (i) of the algorithm in Section 2, let p_i^* indicate the proportion of the bootstrap sample equal to x_i,

$$\text{(10.1)} \qquad p_i^* = \frac{\#\{x_j^* = x_i\}}{n} \quad i = 1, 2, \cdots, n,$$

$\mathbf{p}^* = (p_1^*, p_2^*, \cdots, p_n^*)$. The vector $\mathbf{p}^*$ has a rescaled multinomial distribution

$$\text{(10.2)} \qquad \mathbf{p}^* \sim \mathrm{Mult}_n(n, \mathbf{p}^0)/n \qquad (\mathbf{p}^0 = (1/n, 1/n, \cdots, 1/n)),$$

where the notation indicates the proportions observed from n random draws on n categories, each with probability $1/n$.

For $n = 3$ there are 10 possible bootstrap vectors $\mathbf{p}^*$. These are indicated in Fig. 15 along with their multinomial probabilities from (10.2). For example, $\mathbf{p}^* = (\frac{1}{3}, 0, \frac{2}{3})$, corresponding to $\mathbf{x}^* = (x_1, x_3, x_3)$ or any permutation of these values has bootstrap probability $\frac{1}{9}$.

To make our discussion easier suppose that the statistic of interest $\hat{\theta}$ is of functional form: $\hat{\theta} = \theta(\hat{F})$, where $\theta(F)$ is a functional assigning a real number to any distribution F on the sample space X. The mean, the correlation coefficient, and the trimmed mean are all of functional form. Statistics of functional form have the same value as a function of $\hat{F}$, no matter what the sample size n may be, which is convenient for discussing the jackknife and delta method.

For any vector $\mathbf{p} = (p_1, p_2, \cdots, p_n)$ having nonnegative weights summing to 1, define the weighted

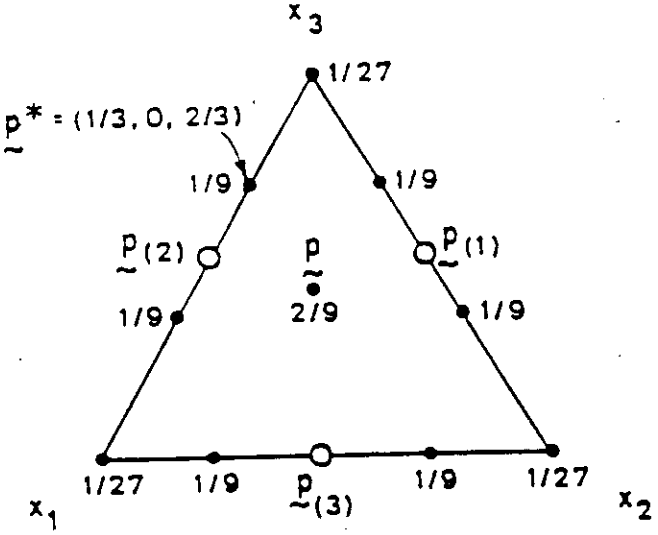


FIG. 15. *The bootstrap and jackknife sampling points in the case $n = 3$. The bootstrap points (·) are shown with their probabilities.*

empirical distribution

$$\hat{F}(\mathbf{p})\text{: probability } p_i \text{ on } x_i \quad i = 1, \cdots, n. \tag{10.3}$$

For $\mathbf{p} = \mathbf{p}^0 = 1/n$, the weighted empirical distribution equals $\hat{F}$, (1.4).

Corresponding to $\mathbf{p}$ is a resampled value of $\hat{\theta}$,

$$\hat{\theta}(\mathbf{p}) \equiv \theta(\hat{F}(\mathbf{p})). \tag{10.4}$$

The shortened notation $\hat{\theta}(\mathbf{p})$ assumes that the data $(x_1, x_2, \cdots, x_n)$ is considered fixed. Notice that $\hat{\theta}(\mathbf{p}^0) = \theta(\hat{F})$ is the observed value of the statistic of interest. The bootstrap estimate $\hat{\sigma}$, (2.3), can then be written

$$\hat{\sigma} = [\operatorname{var}_* \hat{\theta}(\mathbf{p}^*)]^{1/2}, \tag{10.5}$$

where var_* indicates variance with respect to distribution (10.2). In terms of Fig. 15, $\hat{\sigma}$ is the standard deviation of the ten possible bootstrap values $\hat{\theta}(\mathbf{p}^*)$ weighted as shown.

It looks like we could always calculate $\hat{\sigma}$ simply by doing a finite sum. Unfortunately, the number of bootstrap points is $\binom{2n-1}{n}$, 77,558,710 for $n = 15$ so straightforward calculation of $\hat{\sigma}$ is usually impractical. That is why we have emphasized Monte Carlo approximations to $\hat{\sigma}$. Therneau (1983) considers the question of methods more efficient than pure Monte Carlo, but at present there is no generally better method available.

However, there is another approach to approximating (10.5). We can replace the usually complicated function $\hat{\theta}(\mathbf{p})$ by an approximation linear in $\mathbf{p}$, and then use the well known formula for the multinomial variance of a linear function. The *jackknife approximation* $\hat{\theta}_J(\mathbf{p})$ is the linear function of $\mathbf{p}$ which matches $\hat{\theta}(\mathbf{p})$, (10.4), at the n points corresponding to the deletion of a single $\mathbf{x}_i$ from the observed data set $x_1, x_2, \cdots, x_n$,

$$\mathbf{p}_{(i)} = \frac{1}{n-1}(1, 1, \cdots, 1, 0, 1, \cdots, 1) \tag{10.6}$$

$i = 1, 2, \cdots, n$. Fig. 15 indicates the jackknife points for $n = 3$; because $\hat{\theta}$ is the functional form, (10.4), it does not matter that the jackknife points correspond to sample size $n - 1$ rather than n.

The linear function $\hat{\theta}_J(\mathbf{p})$ is calculated to be

$$\hat{\theta}_J(\mathbf{p}) = \hat{\theta}_{(i)} + (\mathbf{p} - \mathbf{p}^0) \cdot \mathbf{U} \tag{10.7}$$

where, in terms of $\hat{\theta}_{(i)} \equiv \hat{\theta}(\mathbf{p}_{(i)})$, $\hat{\theta}_{(\cdot)} = \sum_{i=1}^n \hat{\theta}_{(i)}/n$, and $\mathbf{U}$ is the vector with ith coordinate

$$U_i = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}). \tag{10.8}$$

The jackknife estimate of standard error (Tukey, 1958; Miller, 1974) is

$$\hat{\sigma}_J \equiv \left[\frac{n-1}{n} \sum_{i=1}^n \{\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}\}^2\right]^{1/2} = \left[\frac{\sum_1^n U_i^2}{n(n-1)}\right]^{1/2}. \tag{10.9}$$

A standard multinomial calculation gives the following theorem (Efron, 1982a),

THEOREM. *The jackknife estimate of standard error equals $[n/(n-1)]^{1/2}$ times the bootstrap estimate of standard error for $\hat{\theta}_J$,*

$$\hat{\sigma}_J = \left[\frac{n}{n-1} \operatorname{var}_* \hat{\theta}_J(\mathbf{p}^*)\right]^{1/2}. \tag{10.10}$$

In other words, the jackknife estimate is itself almost a bootstrap estimate applied to a linear approximation of $\hat{\theta}$. The factor $[n/(n-1)]^{1/2}$ in (10.10) makes $\hat{\sigma}_J^2$ unbiased for σ^2 in the case where $\hat{\theta} = \bar{x}$, the sample mean. We could multiply the bootstrap estimate $\hat{\sigma}$ by this same factor, and achieve the same unbiasedness, but there does not seem to be any consistent advantage to doing so. The jackknife requires n, rather than $B = 50$ to 200 resamples, at the expense of adding a linear approximation to the standard error estimate. Tables 1 and 2 indicate that there is some estimating efficiency lost in making this approximation. For statistics like the sample median which are difficult to approximate linearly, the jackknife is useless (see Section 3.4 of Efron, 1982a).

There is a more obvious linear approximation to $\hat{\theta}(\mathbf{p})$ than $\hat{\theta}_J(\mathbf{p})$. Why not use the first-order Taylor series expansion for $\hat{\theta}(\mathbf{p})$ about the point $\mathbf{p} = \mathbf{p}^0$? This is the idea of Jaeckel's *infinitesimal jackknife* (1972). The Taylor series approximation turns out to be

$$\hat{\theta}_T(\mathbf{p}) = \hat{\theta}(\mathbf{p}^0) + (\mathbf{p} - \mathbf{p}^0)'\mathbf{U}^0$$

where

$$U_i^0 = \lim_{\epsilon \to 0} \frac{\hat{\theta}((1-\epsilon)\mathbf{p}^0 + \epsilon\delta_i) - \hat{\theta}(\mathbf{p}^0)}{\epsilon}, \tag{10.11}$$

δ_i being the ith coordinate vector. This suggests the infinitesimal jackknife estimate of standard error

$$\hat{\sigma}_{IJ} = [\mathrm{var}_* \hat{\theta}_T(\mathbf{p}^*)]^{1/2} = [\Sigma U_i^{02}/n^2]^{1/2} \tag{10.12}$$

with var_* still indicating variance under (10.2). The ordinary jackknife can be thought of as taking $\varepsilon = -1/(n-1)$ in the definition of U_i^0, while the infinitesimal jackknife lets $\varepsilon \longrightarrow 0$, thereby earning the name.

The U_i^0 are values of what Mallows (1974) calls the empirical influence function. Their definition is a nonparametric estimate of the true influence function

$$IF(x) = \lim_{\varepsilon \to 0} \frac{\theta((1-\varepsilon)F + \varepsilon\delta_x) - \theta(F)}{\varepsilon},$$

[illegible] the degenerate distribution putting mass 1 on x. The right side of (10.12) is then the obvious estimate of the influence function approximation to the standard error of $\hat{\theta}$ (Hampel, 1974), $\sigma(F) \doteq [\int IF^2(x)\, dF(x)/n]^{1/2}$. The empirical influence function method and the infinitesimal jackknife give identical estimates of standard error.

How have statisticians gotten along for so many years without methods like the jackknife and the bootstrap? The answer is the delta method, which is still the most commonly used device for approximating standard errors. The method applies to statistics of the form $t(\bar{Q}_1, \bar{Q}_2, \cdots, \bar{Q}_A)$, where $t(\cdot, \cdot, \cdots, \cdot)$ is a known function and each $\bar{Q}_a$ is an observed average, $\bar{Q}_a = \sum_{i=1}^n Q_a(X_i)/n$. For example, the correlation $\hat{\theta}$ is a function of $A = 5$ such averages; the average of the first coordinate values, the second coordinates, the first coordinates squared, the second coordinates squared, and the cross-products.

In its nonparametric formulation, the delta method works by (a) expanding t in a linear Taylor series about the expectations of the $\bar{Q}_a$; (b) evaluating the standard error of the Taylor series using the usual expressions for variances and covariances of averages; and (c) substituting $\gamma(\hat{F})$ for any unknown quantity $\gamma(F)$ occurring in (b). For example, the nonparametric delta method estimates the standard error of the correlation $\hat{\theta}$ by

$$\left\{\frac{\hat{\theta}^2}{4n}\left[\frac{\hat{\mu}_{40}}{\hat{\mu}_{20}^2} + \frac{\hat{\mu}_{04}}{\hat{\mu}_{02}^2} + \frac{2\hat{\mu}_{22}}{\hat{\mu}_{20}\hat{\mu}_{02}} + \frac{4\hat{\mu}_{22}}{\hat{\mu}_{11}^2} - \frac{4\hat{\mu}_{31}}{\hat{\mu}_{11}\hat{\mu}_{02}} + \frac{4\hat{\mu}_{13}}{\hat{\mu}_{11}\hat{\mu}_{02}}\right]\right\}^{1/2}$$

where, in terms of $x_i = (y_i, z_i)$,

$$\hat{\mu}_{gh} = \Sigma(y_i - \bar{y})^g(z_i - \bar{z})^h/n$$

(Cramér (1946), p. 359).

THEOREM. *For statistics of the form $\hat{\theta} = t(\bar{Q}_1, \cdots, \bar{Q}_A)$, the nonparametric delta method and the infinitesimal jackknife give the same estimate of standard error* (Efron, 1982c).

The infinitesimal jackknife, the delta method, and the empirical influence function approach are three names for the same method. Notice that the results reported in line 7 of Table 2 show a severe downward bias. Efron and Stein (1981) show that the ordinary jackknife is always biased upward, in a sense made precise in that paper. In the authors' opinion the ordinary jackknife is the method of choice if one does not want to do the bootstrap computations.

ACKNOWLEDGMENT

This paper is based on a previous review article appearing in *Behaviormetrika*. The authors and Editor are grateful to that journal for graciously allowing this revision to appear here.

REFERENCES

ANDERSON, O. D. (1975). *Time Series Analysis and Forecasting: The Box–Jenkins Approach*. Butterworth, London.

BAHADUR, R. and SAVAGE, L. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Statist.* **27** 1115–1122.

BERAN, R. (1984). Bootstrap methods in statistics. *Jahrb. Math. Ver.* **86** 14–30.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations. *J. R. Statist. Soc. Ser. B* **26** 211–252.

BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.* **80** 580–619.

COX, D. R. (1972). Regression models and life tables. *J. R. Statist. Soc. Ser. B* **34** 187–202.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton, New Jersey.

EFRON, B. (1979a). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** 1–26.

EFRON, B. (1979b). Computers and the theory of statistics: thinking the unthinkable. *Soc. Ind. Appl. Math.* **21** 460–480.

EFRON, B. (1981a). Censored data and the bootstrap. *J. Amer. Statist. Assoc.* **76** 312–319.

EFRON, B. (1981b). Nonparametric estimates of standard error: the jackknife, the bootstrap, and other resampling methods. *Biometrika* **68** 589–599.

EFRON, B. (1982a). The jackknife, the bootstrap, and other resampling plans. *Soc. Ind. Appl. Math. CBMS-Natl. Sci. Found. Monogr.* **38**.

EFRON, B. (1982b). Transformation theory: how normal is a one parameter family of distributions? *Ann. Statist.* **10** 323–339.

EFRON, B. (1982c). Maximum likelihood and decision theory. *Ann. Statist.* **10** 340–356.

EFRON, B. (1983). Estimating the error rate of a prediction rule: improvements in cross-validation. *J. Amer. Statist. Assoc.* **78** 316–331.

EFRON, B. (1984). Better bootstrap confidence intervals. Tech. Rep. Stanford Univ. Dept. Statist.

EFRON, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika* **72** 45–58.

EFRON, B. and GONG, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *Amer. Statistician* **37** 36–48.

EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596.

FIELLER, E. C. (1954). Some problems in interval estimation. *J. R. Statist. Soc. Ser. B* **16** 175–183.

FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823.

FRIEDMAN, J. H. and TIBSHIRANI, R. J. (1984). The monotone smoothing of scatter-plots. *Technometrics* 26 243–250.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* 69 383–393.

HASTIE, T. J. and TIBSHIRANI, R. J. (1985). Discussion of Peter Huber's "Projection Pursuit." *Ann. Statist.* 13 502–508.

HINKLEY, D. V. (1978). Improving the jackknife with special reference to correlation estimation. *Biometrika* 65, 13–22.

HYDE, J. (1980). *Survival Analysis with Incomplete Observations. Biostatistics Casebook.* Wiley, New York.

JAECKEL, L. (1972). The infinitesimal jackknife. Memorandum MM 72-1215-11. Bell Laboratories, Murray Hill, New Jersey.

JOHNSON, N. and KOTZ, S. (1970). *Continuous Univariate Distributions.* Houghton Mifflin, Boston, Vol. 2.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete samples. *J. Amer. Statist. Assoc.* 53 457–481.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* 27 887–906.

MALLOWS, C. (1974). On some topics in robustness. Memorandum. Bell Laboratories, Murray Hill, New Jersey.

MILLER, R. G. (1974). The jackknife—a review. *Biometrika* 61 1–17.

MILLER, R. G. and HALPERN, J. (1982). Regression with censored data. *Biometrika* 69 521–531.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York.

SCHENKER, N. (1985). Qualms about bootstrap confidence intervals. *J. Amer. Statist. Assoc.* 80 360–361.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* 9 1187–1195.

THERNEAU, T. (1983). Variance reduction techniques for the bootstrap. Ph.D. thesis, Stanford University, Department of Statistics.

TIBSHIRANI, R. J. and HASTIE, T. J. (1984). Local likelihood estimation. Tech. Rep. Stanford Univ. Dept. Statist. 97.

TUKEY, J. (1958). Bias and confidence in not quite large samples, abstract. *Ann. Math. Statist.* 29 614.

Comment

J. A. Hartigan

Efron and Tibshirani are to be congratulated on a wide-ranging persuasive survey of the many uses of the boostrap technology. They are a bit cagey on what is or is not a bootstrap, but the description at the end of Section 4 seems to cover all the cases; some data y comes from an unknown probability distribution F; it is desired to estimate the distribution of some function $R(y, F)$ given F; and this is done by estimating the distribution of $R(y^*, \hat{F})$ given $\hat{F}$ where $\hat{F}$ is an estimate of F based on y, and y^* is sampled from the known $\hat{F}$.

There will be three problems in any application of the bootstrap: (1) how to choose the estimate $\hat{F}$? (2) how much sampling of y^* from $\hat{F}$? and (3) how close is the distribution of $R(y^*, \hat{F})$ given $\hat{F}$ to $R(y, F)$ given F?

Efron and Tibshirani suggest a variety of estimates $\hat{F}$ for simple random sampling, regression, and autoregression; their remarks about (3) are confined mainly to empirical demonstrations of the bootstrap in specific situations.

I have some general reservations about the bootstrap based on my experiences with subsampling techniques (Hartigan, 1969, 1975). Let $X_1, \ldots, X_n$ be a random sample from a distribution F, let F_n be the empirical distribution, and suppose that $t(F_n)$ is an estimate of some population parameter $t(F)$. The statistic $t(\hat{F}_n)$ is computed for several random subsamples (each observation appearing in the subsample with probability ½), and the set of $t(\hat{F}_n)$ values obtained is regarded as a sample from the posterior distribution of $t(F)$. For example, the standard deviation of the $t(\hat{F}_n)$ is an estimate of the standard error of $t(F_n)$ from $t(F)$; however, the procedure is not restricted to real valued t.

The procedure seems to work not too badly in getting at the first- and second-order behaviors of $t(F_n)$ when $t(F_n)$ is near normal, but it not effective in handling third-order behavior, bias, and skewness. Thus there is not much point in taking huge samples $t(\hat{F}_n)$ since the third-order behavior is not relevant; and if the procedure works only for $t(F_n)$ near normal, there are less fancy procedures for estimating standard error such as dividing the sample up into 10 subsamples of equal size and computing their standard deviation. (True, this introduces more bias than having random subsamples each containing about half the observations.) Indeed, even if $t(F_n)$ is not normal, we can obtain exact confidence intervals for the median of $t(F_{n/10})$ using the 10 subsamples. Even five subsamples will give a respectable idea of the standard error.

Transferring back to the bootstrap: (A) is the boot-

J. A. Hartigan is Eugene Higgins Professor of Statistics, Yale University, Box 2179 Yale Station, New Haven, CT 06520.

strap effective for non-normal situations? (B) in the normal case, does the bootstrap give accurate assessment of third-order terms? If not, it is scarcely justified to do many bootstrap simulations, since you will only use them to estimate a variance. The asymptotic justifications of the bootstrap such as in Bickel and Freeman (1981) or Singh (1981) do consider behavior near the normal.

To be specific, consider the case where a statistic $t(F_n)$ estimates a parameter $t(F)$. The first kind of bootstrapping might be on the quantity $t(F_n) - t(F)$; to estimate its variance $\sigma^2(F)/n$ we compute repeatedly $t(\hat{F}_n) - t(F_n)$ where $\hat{F}_n$ is the empirical distribution of a sample of size n from F_n. Thus $\sigma^2(F_n)$ will be used to estimate $\sigma^2(F)$. We might hope that

$$t(F_n) = t(F) + \xi \frac{\sigma(F)}{\sqrt{n}} + O\left(\frac{1}{n}\right)$$

where $\xi \sim N(0, 1)$. This is the case referred to above where $t(F_n)$ is normal and numerous resampling estimates are available to estimate $\sigma^2(F)$. To do better, consider the higher order terms:

$$t(F_n) = t(F) + \xi \frac{\sigma(F)}{\sqrt{n}} + \frac{s_3(F)}{n} (\xi^2 - 1) + \frac{b(F)}{n} + O(n^{-3/2}).$$

Then

$$t(\hat{F}_n) = t(F_n) + \xi \frac{\sigma(F_n)}{\sqrt{n}} + \frac{s_3(F_n)}{n} (\xi^2 - 1) + \frac{b(F_n)}{n} + O(n^{-3/2}).$$

We might expect that the sample quantities $\sigma(F_n)$, $s_3(F_n)$; $b(F_n)$ are within $O(n^{-1/2})$ of the population quantities; but since $\sigma(F_n) - \sigma(F) = O(n^{-1/2})$, the error in approximating the distribution of $t(F_n) - t(F)$ by that of $t(\hat{F}_n) - t(F_n)$ is $O(n^{-1/2})$, so that the additional skewness and bias terms are of no interest:

$$P\left[(t(F_n) - t(F) \leq \frac{a}{\sqrt{n}}\right] - P\left[t(\hat{F}_n) - t(F_n) \leq \frac{a}{\sqrt{n}}\right] = O(n^{-1/2}).$$

The bootstrap distribution is no better than any normal approximation using an estimate of variance accurate to $O(n^{-1/2})$!

On the other hand, if

$$R(y, F) = [t(F_n) - t(F)]/\sigma(F),$$

$$[t(F_n) - t(F)]/\sigma(F) = \frac{\xi}{\sqrt{n}} + \frac{s_3'(F)}{n} (\xi^2 - 1) + \frac{b'(F)}{n} + O(n^{-3/2})$$

$$[t(\hat{F}_n) - t(F_n)]/\sigma(F_n) = \frac{\xi}{\sqrt{n}} + \frac{s_3'(F_n)}{n} (\xi^2 - 1) + \frac{b'(F_n)}{n} + O(n^{-3/2}).$$

Now $s_3'(F_n)$ estimates $s_3'(F)$ and $b'(F_n)$ estimates $b'(F)$ to within $O(n^{-1/2})$, and the Cornish-Fisher expansion is accurate to skewness and bias terms:

$$P\left(\frac{t(F_n) - t(F)}{\sigma(F)} \leq \frac{a}{\sqrt{n}}\right) - P\left(\frac{t(\hat{F}_n) - t(F_n)}{\sigma(F_n)} \leq \frac{a}{\sqrt{n}}\right) = O(n^{-1}).$$

These results are given for $t(F_n) = \bar{X}$ in Singh (1981).

The conclusion is that for $t(F_n)$ near normal there is no advantage for the bootstrap over other resampling methods, unless the pivotal $[t(F_n) - t(F)]/\sigma(F)$ is used. Usually $\sigma(F)$ is not known; that's why we are resampling in the first place. We would need to estimate it by bootstrapping and use the pivotal $(t(F_n) - t(F))/\sigma(F_n)$. And the distribution of this pivotal would be determined by bootstrapping to obtain $[t(\hat{F}_n) - t(F_n)]/\sigma(\hat{F}_n)$. Note that $\sigma(\hat{F}_n)$ requires two levels of bootstrapping; this might get close to Professor Efron's objective of soaking up all the spare cycles on the West Coast!

Let us consider the modest objective of estimating the variance of $t(F_n)$. The various resampling techniques compute the variance of $t(W^1)$, $t(W^2)$, ..., $t(W^k)$ where $t(W^j)$ denotes the statistic computed on X_i repeated W_i^j times, $1 \leq i \leq n$. What is a good choice of $W^1, W^2, \dots, W^k$? If in fact $X_1, \dots, X_n$ are sampled from $N(\mu, \sigma^2)$ and $t = \bar{X}$, a minimum variance unbiased estimate of σ^2 is obtained by setting $W_i^j = 1 + \sqrt{n}\xi_i^j$ where ξ^1, σ^2, ..., ξ^k are any k orthonormal vectors orthogonal to 1. The quantities $\sqrt{n}\xi_i^j$ can be obtained roughly by sampling each of them independently from $N(0, 1)$. Bootstrap resampling, for large n, has W_i^j approximately independently Poisson with expectation 1. Random subsampling, for large n has W_i^j approximately independent and approximately taking values 0 and 2 with probability ½. The Dirichlet distribution for F given F_n produces weights W_i^j that are approximately exponential with expectation 1. Any resampling scheme in which the weights are approximately independent with mean and variance 1 will give the right expected variance, but the efficiency of the estimate (at normal means) is optimal for $W_i' = 1 + \sqrt{n}\xi_i'$.

For $n = 8$, obtain an efficient estimate from subsamples (1234), (1256), (1278), (1357), (1368), (1458), (1467); use as many as you need, and if $n > 8$ divide the sample into 8 groups as evenly as possible. I think it must be rare that the various approximations needed to connect the resampled computation to the computation of interest will be satisfied well enough to justify more than a few resamples. Perhaps this method might be called the *shoestring.*

ADDITIONAL REFERENCES

Hartigan, J. A. (1969). Using subsample values as typical values. *J. Amer. Statist. Assoc.* 64 1303–1317.

Hartigan, J. A. (1975). Necessary and sufficient conditions for asymptotic joint normality of a statistic and its subsample values. *Ann. Statist.* 3 573–580.

Rejoinder

B. Efron and R. Tibshirani

Professor Hartigan, who is one of the pioneers of resampling theory, raises the question of higher order accuracy. This question has bothered resamplers since the early days of the jackknife. Sections 7 and 8 of our paper show that the bootstrap can indeed achieve higher levels of accuracy, going the next step beyond simple estimates of standard error. The bootstrap confidence intervals we discuss are *not* of the crude (although useful) first-order form $\hat{\theta} \pm \hat{\sigma} z^{(\alpha)}$. They explicitly incorporate the higher order corrections about which Hartigan is legitimately concerned.

In particular the "z_0" term (7.8) is a correction for bias, and the acceleration constant "a," (7.16), is a correction for skewness. These correspond to Hartigan's $b(F)$ and $s_3(F)$, respectively. The reader who follows through Tables 5 and 7 will see these corrections in action. The fact that they produce highly accurate confidence intervals is no accident. The theory in Efron (1984a, 1984b) demonstrates higher order accuracy of the BC_a intervals in a wide class of situations. This demonstration does not yet apply to fully general problems, but current research indicates that it soon will. (The impressive higher order asymptotic results of Beran, Singh, Bickel, and Freedman, referred to in the paper, underpin these conclusions.)

It is worth mentioning that the bias and skewness corrections of the bootstrap confidence intervals are not of the simple "plug into an approximate pivotal" form suggested in Hartigan's remarks. The theory is phrased in a way which automatically corrects for arbitrary nonlinear transformations, even of the violent sort encountered in the correlation example of Table 5. In this sense the bootstrap theory does handle "non-normal situations."

Since this paper was written, research by several workers, including T. Hesterberg, R. Tibshirani, and T. DiCiccio, has substantially improved the computational outlook for bootstrap confidence intervals. It now appears possible that bootstrap sample sizes closer to $B = 100$ than $B = 1000$ may be sufficient for the task. However, these improvements are still in the process of development.

Professor Hartigan's last remarks, on the comparative efficiency of different resampling methods, need careful interpretation. There are two concepts of efficiency involved: the efficiency of the numerical algorithm in producing an estimate of variance, and the statistical efficiency of the estimate produced. There is no question that other resampling techniques, for example, the jackknife, can produce variance estimates more economically than does the bootstrap. We have argued, both by example and theory, that the bootstrap variance is generally more efficient as a statistical estimator of the unknown true variance.

This is not surprising given that methods like the jackknife are Taylor series approximations to the bootstrap (see Section 10). The simple idea in (2.3), substituting $\hat{F}$ for F, lies at the heart of all nonparametric estimates of accuracy. The bootstrap is the crudest of these methods in that it computes $\sigma(\hat{F})$ directly by Monte Carlo. For this reason it is also the method that involves the least amount of analytic approximation. It is perhaps surprising, and certainly gratifying, that a method based on such a simple form of inference is capable of producing quite accurate confidence intervals.

To say that the bootstrap is good, as we have been blatantly doing, doesn't imply that other methods are bad. Professor Hartigan's own work shows that for some problems, for example, forming a confidence interval for the center of a symmetric distribution, other methods are better. We hope that resampling methods in general will continue to be a lively research topic.

APPENDIX E

EXPOSURE TO ACETALDEHYDE FROM INDOOR AIR

EXPOSURE TO ACETALDEHYDE FROM INDOOR AIR

I. BACKGROUND

California Health and Safety Code Section 39660.5 directs the Board to assess exposures to toxic air contaminants in indoor as well as outdoor environments, and to identify the contribution of indoor concentrations to total air exposure. Indoor exposure assessment has become increasingly important as an integral part of assessing exposure to toxic air contaminants because:

1. Californians spend most of their time (about 87 percent on average) indoors (Wiley et al., 1989); and
2. personal and indoor air monitoring data indicate that some pollutant concentrations are regularly higher indoors than outdoors.

Thus, consideration of indoor air exposure data is critical in developing a realistic estimate of personal exposures through the air environment.

While the main objective of this report is to estimate the exposure to acetaldehyde in the indoor air environment, limited data are also presented on exposure through other media, such as food and water. This is not by any means an exhaustive review of all available data, but is included to provide a broader perspective of environmental exposure to acetaldehyde.

II. INDOOR AIR CONCENTRATIONS OF ACETALDEHYDE

Only a small number of limited surveys that measured acetaldehyde concentrations in indoor air have been published. Those surveys monitored a small number of homes, and the homes that were monitored were not selected randomly. With such limitations, it is not possible to extrapolate the monitoring results to apply to the general population of homes in the individual surveyed. However, when viewed together, the results of individual surveys can at least provide a general idea of the magnitude of indoor acetaldehyde concentrations. It is in that context that those surveys are presented below.

A. Residences

In one of the larger published surveys, Zweidinger et al. (1988) measured acetaldehyde concentrations inside 20 homes of non-smokers in Boise, Idaho from November, 1986, to February 1987. From Saturday through Tuesday during each week of the survey the investigators measured acetaldehyde concentrations inside a pair of homes; each pair consisted of one home with a woodstove and one without. Consecutive 12-hour air samples were obtained from each home using 2,4-dinitrophenylhydrazine (2,4-DNPH) coated silica gel cartridges. Sample analysis was performed by high performance liquid chromatography (HPLC). Average acetaldehyde concentrations (as calculated from data presented) inside the homes ranged from 9.3 to 11.8 ppbv (16.7 to 21.2 $\mu g/m^3$). The presence of a woodstove did not appear to affect indoor acetaldehyde concentrations. Concurrently-sampled average outdoor acetaldehyde concentrations (sampled only outside the homes without woodstoves) ranged from 2.1 to 2.6 ppbv (3.8 to 4.7 $\mu g/m^3$) (as calculated from data presented).

A similar survey was recently conducted in Roanoke, Virginia, the results of which have not yet been published (R. Zweidinger, personal communication). Twenty homes of non-smokers were monitored in pairs that consisted of one home that used oil as a heating fuel and one home that did not. The monitoring methods and sample collection times were the same as those used in the Boise survey. Preliminary data indicate that the average acetaldehyde concentrations in those homes was 8.3 ppbv (14.9 $\mu g/m^3$). Indoor concentrations ranged 3 to 26 ppbv (5.4 to 46.8 $\mu g/m^3$), but contamination of some of the sample blanks casts doubt on the validity of some of the high values. At this preliminary stage of the data analysis, conclusions have not been drawn concerning the effect of oil heating on indoor acetaldehyde concentrations.

In an earlier pilot study conducted in February, 1985 (Zweidinger et al., 1987; Highsmith et al., 1988) the investigators used similar methods to measure acetaldehyde concentrations inside and outside three homes in Raleigh, North Carolina. All three homes were occupied by non-smokers and were monitored during two nights for 12 hours per night (one house was monitored only once) while their woodstoves were in operation. The investigators reported that the average indoor acetaldehyde concentrations reported ranged from 4.0 to 8.5 ppbv (7.2 to 15.3 $\mu g/m^3$). Average concurrently-measured outdoor concentrations ranged from 1.3 to 4.4 ppbv (2.3 to 7.9 $\mu g/m^3$); the highest outdoor value was observed outside a home where the chimney plume down washed directly on top of the outdoor samplers.

Limited measurements of acetaldehyde inside two California homes were performed in 1983 by Rogozen et al. (1984). Those two homes were selected from 64 homes, located throughout the State, which had been randomly selected for a formaldehyde survey. Air inside the two homes was sampled for one hour using impingers containing a 2,4-DNPH solution and analysis was performed by HPLC. The reported acetaldehyde concentrations inside the two homes were 7.1 and 2.6 ppbv (12.8 and 4.7 $\mu g/m^3$).

Acetaldehyde levels were measured in a subsample of 8 homes in Baltimore, Maryland, in 1987 as part of a larger survey of 155 homes (W. Nelson, personal communication; Nelson et al., 1988). In the larger survey, the 155 homes (which included homes of smokers) were selected at random; the 8-home subsample was comprised of homes that were monitored during one week of the larger survey, and were thus not selected strictly at random. Two consecutive 12-hour samples were collected inside and outside each of the 8 homes using 2,4-DNPH-coated silica gel cartridges followed by HPLC analysis. The mean indoor acetaldehyde concentration was 14.8 ppbv (26.6 $\mu g/m^3$) and the values ranged from 6.2 to 37.4 ppbv (11.2 to 67.3 $\mu g/m^3$). Concurrent outdoor acetaldehyde concentrations averaged 1.7 ppbv (3.1 $\mu g/m^3$) and ranged from 0.3 to 4.1 ppbv (0.5 to 7.4 $\mu g/m^3$).

In an unpublished pilot study conducted in the Raleigh/Durham area of North Carolina in 1983, investigators measured acetaldehyde concentrations inside 15 residences (12 conventional homes and 3 mobile homes; S. Tejada, personal communication). Air samples were collected for one to two hours using impingers containing 2,4-DNPH. The mean concentration measured in all 15 homes was 20.4 ppbv (36.7 $\mu g/m^3$); one high value measured in a mobile home drove this mean value up somewhat. The mean acetaldehyde concentration for just the 12 conventional homes was 15.2 ppbv (27.4 $\mu g/m^3$).

Another unpublished pilot study was conducted in the same area in 1984 (J. Bufalini, personal communication). In that study, investigators measured acetaldehyde concentrations inside two homes during times when their fireplaces were in use and not in use. Several 30-minute samples were collected inside each home using impingers containing 2,4-DNPH. The acetaldehyde concentrations ranged from 3.6 to 15.0 ppbv (6.5 to 27.0 $\mu g/m^3$). Fireplace usage did not appear to alter indoor acetaldehyde concentrations significantly.

B. Public Buildings

Bufalini and his co-investigators also used similar methods to measure acetaldehyde concentrations inside two offices located in Research Triangle Park, North Carolina (J. Bufalini, personal communication). The acetaldehyde concentration inside a non-smoking office was 1.2 ppbv (2.2 $\mu g/m^3$), while that inside an office where smoking occurred was 12.8 ppbv (23.0 $\mu g/m^3$). The offices were located down the hall from each other and had similar air exchange rates (approximately 4 to 5 air changes per hour).

As part of a study on environmental tobacco smoke, Lofroth et al. (1989) measured acetaldehyde concentrations inside a North Carolina tavern which was about 180 m^3 in volume and was variously occupied by 5 to 25 people, many of whom were smoking. Samples were collected on two separate days for three and four hours, respectively, using 2,4-DNPH-coated silica gel cartridges followed by HPLC analysis. The reported acetaldehyde concentrations measured for those two days was 102 ppbv; using a conversion factor of 1 ppbv = 1.8 $\mu g/m^3$ (183 $\mu g/m^3$) and 113 ppbv (204 $\mu g/m^3$).

Druzik et al., (1990) measured indoor concentrations of acetaldehyde in five museums and one library located in Los Angeles. Samples were collected with DNPH-coated cartridges. Indoor concentrations ranged from 4.2 to 35 ppbv (7.6 to 63.0 $\mu g/m^3$). Outdoor concentrations measured in Los Angeles using the same method were lower; they ranged from 1.7 to 9.9 ppbv (3.1 to 17.8 $\mu g/m^3$).

C. In-vehicle Concentrations

The air in the enclosed space encountered in vehicles is not strictly "indoor air", but measurements of acetaldehyde concentrations in such enclosed spaces are included here to facilitate consideration of people's total exposures to acetaldehyde. A study which measured acetaldehyde levels in the cars of volunteers during their daily commute to and from an office complex in El Monte, California was conducted from May to October, 1987 and from November to March, 1988. Measurements were made using a cartridge impregnated with 2,4-DNPH with HPLC analysis. The daily one-way commute of the volunteers averaged 33 minutes; the mean acetaldehyde concentrations of 194 total samples was 7.6 ppbv (13.7 $\mu g/m^3$). The maximum concentration was 37.0 ppbv (66.6 $\mu g/m^3$) (Shikiya et al., 1989).

D. Summary of Indoor Concentrations

There is only limited information available concerning indoor concentrations of acetaldehyde. Based on the results of a number of small-scale surveys (summarized in Table 1), a very crude estimate of an average acetaldehyde concentration that would be expected to occur inside residences is about 3 to 15 ppbv (5.4 to 27.0 $\mu g/m^3$). The highest acetaldehyde concentration reported inside a residence was 37.4 ppbv (67.3 $\mu g/m^3$). The results of one study suggest that higher levels may occur in some indoor environments, in this case a tavern occupied by a number of people who were smoking, where levels of over 100 ppbv (180 $\mu g/m^3$) were reported. Average and maximum in-vehicle acetaldehyde concentrations [7.6 and 37.0 ppbv (13.7 and 66.6 $\mu g/m^3$), respectively] appear to be similar in magnitude to those inside residences.

III. POTENTIAL SOURCES OF ACETALDEHYDE IN INDOOR AIR

Acetaldehyde is formed as a combustion by-product and can be emitted from a number of indoor sources including cigarettes, fireplaces and woodstoves, and cooking. Acetaldehyde occurs in traces in all ripe fruit and may form in alcoholic beverages after exposures to the air (Fishbein, 1979), although it is not known if volatilization from those sources could affect indoor acetaldehyde concentrations to any significant degree. Vehicle exhaust may also contribute to indoor acetaldehyde levels. Limited information indicates that acetaldehyde is emitted from some building materials and consumer products.

A. Cigarette Smoke

Acetaldehyde has been reported in fresh leaf tobacco as well as in tobacco smoke (Fishbein, 1979). Mainstream cigarette smoke (the smoke that

TABLE 1

Indoor Air Concentrations of Acetaldehyde*
(parts per billion)

LOCATION	AVERAGE CONCENTRATION	RANGE	N	COMMENTS	REFERENCES
<u>RESIDENCES</u> Boise, ID	9.3-11.8		20	non-smokers	Zweidinger et al., 1988
California		2.6 & 7.1	2		Rogozen et al., 1984
Raleigh, NC	4.0-8.5		3	non-smokers	Zweidinger, et al., 1987 Highsmith, et al., 1988
Roanoke, VA	8.3	3-26**	20	non-smokers	Zweidinger, personal communication
Baltimore, MD	14.8	6.2-37.4	8	smokers may be included	Nelson, personal communication
Raleigh/ Durham, NC	20.4 15.2		15 12	+3 mobile -3 mobile	Tejada, personal communication
Raleigh/ Durham, NC		3.6-15	2		Bufalini, personal communication
<u>PUBLIC BLDGS.</u> Los Angeles, CA 5 museums, 1 library		4.2-35	6		Druzik et al., 1990
North Carolina office bldg.		1.2 12.8	1 1	non-smoking smoking	Bufalini, personal communication
North Carolina tavern		102 & 113	1	heavy smoking	Lofroth et al., 1989

(continued on next page)

INDOOR AIR CONCENTRATIONS OF ACETALDEHYDE continued

LOCATION	AVERAGE CONCENTRATION	RANGE	N	COMMENTS	REFERENCES
<u>INSIDE VEHICLES</u>					
Southern California	7.6	max. 37.0	194	commute driving	Shikiya et al., 1989
<u>SMOKING CHAMBER</u>					
		48 & 52		1 cig./30 min.	Lofroth et al., 1989
		114 & 129		1 cig./15 min.	

* Data on indoor air concentrations are limited; data cannot be extrapolated to apply to general population of homes in the respective areas surveyed. The data presented here are intended to provide a general idea of the magnitude of indoor acetaldehyde concentrations.

** Contamination of some field blanks casts doubt on the validity of some of the high values.

is generated during a puff and is inhaled by the smoker) contains from 400 to 1,400 μg of acetaldehyde per unfiltered cigarette (U. S. Department of Health and Human Services, 1989). Lofroth et al. (1989) recently measured acetaldehyde concentrations in sidestream cigarette smoke (smoke which arises mainly from the passive burning of the cigarette and is released into the environment) and determined the airborne yield per cigarette to be 2,400 μg. It should be noted that the above values were determined using machine-smoked cigarettes under standard laboratory conditions, which do not reflect present-day smoking behavior (U.S. Department of Health and Human Services, 1989).

B. Woodsmoke and Cooking

Acetaldehyde is formed as a product of incomplete wood combustion in residential fireplaces and woodstoves. It is among the major aldehydes emitted from wood-burning fireplaces (Lipari et al., 1984). Although the results of limited residential surveys suggest that properly-operating fireplaces and woodstoves do not affect indoor acetaldehyde concentrations to any significant degree (see section II.), wood smoke can enter the living space when wind causes variations in room pressure or when flues or chimneys are blocked. Smoke which exits the residence or smoke from nearby neighbors might also be drawn indoors under certain meteorological conditions (Quraishi and Todd, 1987).

Sexton et al. (1986) noted that acetaldehyde was a major constituent of emissions from frying hamburger (levels not specified).

C. Other Indoor Sources

Vehicle exhausts. Vehicle exhaust contains acetaldehyde (Rogozen et al., 1987) and could contribute to indoor concentrations. For example, vehicle exhaust from attached garages or a nearby busy roadway may enter residences.

Building materials. Acetaldehyde emissions were detected from four sprayed-in-place rigid polyurethane foams (Krzymien, 1989). Such foams are used as insulation in aircraft, recreational vehicles, commercial buildings, and, to a limited extent, in residences.

Consumer products. Small amounts of acetaldehyde may be present as a contaminant in acetone, which is widely used in nail polish remover and other consumer products (Rogozen et al., 1987). Sexton et al. (1986) noted that acetaldehyde was a major constituent in the emissions from a deodorant spray (levels not specified). Acetaldehyde may be emitted from diverse products such as lubricants, inks, adhesives and coatings (Hodgson and Wooley, 1991).

IV. OTHER ROUTES OF EXPOSURE TO ACETALDEHYDE

Limited information is presented on exposure through routes other than inhalation. This is not an exhaustive review of all available data, but rather has been included to provide a broader perspective of environmental exposure to acetaldehyde.

A. Ingestion

Food. Acetaldehyde is a normal intermediate product in the respiration of higher plants (Fishbein, 1979) and is thus found naturally in many fruits. Acetaldehyde has been detected in apples, broccoli, coffee, grapefruit, strawberries, and blueberries (WHO, 1985). Levels measured in orange and grapefruit juices and other citrus products ranged from 50 to 190 ppm (w/v) (Lund et al., 1981). Acetaldehyde is found in the essential oils of various natural flavorings and scents. It has been detected in cheese, heated skim milk, cooked beef, cooked chicken, and rum. It has also been detected at trace levels in two kinds of mushrooms (WHO, 1985).

Acetaldehyde is also added to foods as a flavoring agent and a preservative. It is "generally recognized as safe" by the U.S. Food and Drug Administration for use as a synthetic flavoring substance and adjuvant (WHO, 1985). The FAO/WHO (Food and Agriculture Organization/World Health Organization) has listed the acceptable daily intake of synthetic flavorings (which specifically includes acetaldehyde) at 0 to 2.5 mg/kg body weight, with a typical level of use listed at 1-300 ppm (Doull et al., 1980). Acetaldehyde is an important component of many flavors added to foods, such as milk products, baked goods, fruit juices, candies, desserts, and soft drinks, at usual levels of up to 0.047 percent. It is present in flavoring used to impart a butter-like flavor to processed foods, especially margarine. Acetaldehyde is also used as a preservative for fruit and fish and is used as a storage fumigant for foods such as apples and strawberries (WHO, 1985).

Acetaldehyde is a product of alcohol fermentation and has been measured in 18 European beers at concentrations ranging from 2.6 to 13.5 mg/l (WHO, 1985). Additionally, acetaldehyde may form in alcoholic beverages after exposure to the air (Fishbein, 1979). Alcohol is a major source of acetaldehyde in the body; acetaldehyde is a metabolic intermediate in the oxidation of ethanol by liver enzymes (NRC, 1981). It should be noted that acetaldehyde is an intermediate product in the metabolism of sugars in the body and normally occurs at trace levels in the blood (Fishbein, 1979).

Water. Acetaldehyde levels were recently measured in ten California water treatment facilities located throughout the State. The facilities selected use source waters that are representative of supplies used by the majority of consumers in California. The results show that the median acetaldehyde concentration in drinking water samples was 2.1 μg/l and the 75th percentile concentration was approximately four μg/l (Metropolitan Water District of Southern California and JMM, Inc., 1989). Thus, a person drinking two liters of water per day may receive up to about 8 μg of acetaldehyde per day through this route.

B. Skin Absorption

No data on dermal absorption of acetaldehyde are readily available. However, based on its chemical similarity to formaldehyde, which is not absorbed systemically through the dermal route to any appreciable degree (World Health Organization, 1987, 29 CFR Parts 1910 and 1926), dermal absorption of acetaldehyde is likely to be negligible.

C. Summary of Other Routes of Exposure

Acetaldehyde is found commonly in a number of foods as a natural constituent and as an intentional food additive. The daily amount ingested through food cannot be estimated using readily available information. Levels in water are relatively low; a high estimate of the amount a Californian may ingest through water is 8 μg/day. Dermal absorption is surmised to be negligible. It should be noted that, with respect to the potential carcinogenicity of acetaldehyde, the ingestion route may not be of major significance; it is surmised that only in specialized organs such as the nasal cavity that inhalation of appreciable amounts may overcome local defense mechanisms (Casarett et al., 1986).

V. ESTIMATES OF THE PRESENTED DOSE OF ACETALDEHYDE FROM RESIDENTIAL INDOOR AIR AND ENVIRONMENTAL TOBACCO SMOKE

The inhalation route is of major importance to the toxic (eg. carcinogenic) effects of acetaldehyde. Estimates of the daily amounts of acetaldehyde inhaled through residential indoor air and from environmental tobacco smoke are shown in Table 2. These estimates are based on very limited data and are thus very crude. The term "presented dose" is used because the presented dose may not necessarily represent a true dose, since it does not take into account the fraction of the inhaled amount that is absorbed so that the already high level of uncertainty in the values reported is not increased further.

The estimates of presented doses from residential exposures were calculated using the indoor concentration estimates derived in section II and an average value for breathing volume. Those estimates assume that an average person inhales about the same indoor concentration of acetaldehyde (i.e the residential concentration) all day. This assumption was made because: 1) Californians spend most of their time indoors in the home (Wiley et al., 1989); and 2) limited evidence suggests that acetaldehyde

TABLE 2

Estimates of the Presented Dose of Acetaldehyde From Residential Indoor Air and Environmental Tobacco Smoke

	Acetaldehyde inhaled (micrograms per day)
Residential indoor air*	
a. average amount inhaled	108 - 540
b. high range of estimate	1,440
Environmental tobacco smoke**	1,740 - 4,660

* Based on very limited data; office building exposures and in-vehicle exposures are estimated to be very similar to residential exposures.

** Based on chamber and tavern concentrations measured by Lofroth et al. (1989) under specific conditions; does not necessarily represent range of all possible levels.

Assumptions:

1. The average person inhales about 20 cubic meters of air per day.
2. Ingestion and skin absorption are not significant routes of exposure.
3. A person inhales about the same acetaldehyde concentration all day.

concentrations inside vehicles and offices may be similar in magnitude to residential levels. Because the indoor concentration data are so limited, there would be little benefit in attempting to refine the estimates further (eg. by taking into account the average amount of time per day Californians spend in the home), so no such refinements were attempted.

Based on limited residential surveys, the average residential indoor acetaldehyde concentration is 3 to 15 ppbv (5.4 to $27.0\ \mu g/m^3$) and may range up to approximately 40 ppbv ($72\ \mu g/m^3$). The average adult inhales approximately 20 m^3 of air per day (International Commission on Radiological Protection, 1975). Thus, a person may inhale approximately 108 to 540 μg of acetaldehyde per day on average. The amount inhaled could potentially range up to 1,440 μg/day.

A crude estimate of the amount of acetaldehyde that could potentially be inhaled from environmental tobacco smoke was calculated based on acetaldehyde concentrations measured in a smoking chamber and in a tavern with smokers present (see sections II and III). For simplicity, and because the data upon which the estimate is based are limited, the estimate assumes that the acetaldehyde concentration inhaled is about the same throughout the day. Acetaldehyde concentrations produced by sidestream smoke in a smoking chamber ranged from 87 $\mu g/m^3$ (one cigarette smoked every 30 minutes) to 233 $\mu g/m^3$ (one cigarette smoked every 15 minutes). A person breathing 20 m^3 of air would thus inhale between 1,740 and 4,660 μg of acetaldehyde concentration of III.) would inhale 4,080 μg of acetaldehyde per day. This value is within the range estimated using the chamber data.

Summary. Individuals are estimated to inhale about 108-540 μg of acetaldehyde per day in indoor residential environments and other indoor environments, on the average. Daily inhaled doses from indoor air may range up to 1,440 μg. If an individual were to spend the entire day in a smoke-filled environment such as a tavern, the daily inhaled dose could total from 1,740 to 4,660 μg. These estimates were based on very limited data and may not be representative of the California population.

Bufalini, J. (1990) Environmental Protection Agency, Atmospheric Sciences Research Laboratory, North Carolina, personal communication.

Casarett, L., Doull, J., Klaassen, C., Amdur, M., eds. (1986) Casarett and Doull's Toxicology: The Basic Science of Poisons, third edition, Macmillan Publishing Company, New York, pp 598-601.

Doull, J., Klaassen, C. (1980) M. Amdur, eds. Casarett and Doull's Toxicology: The Basic Science of Poisons, second edition, Macmillan Publishing Company, New York, pp 598-601.

Druzik, C., Grosjean, D., Van Neste, A., Parmar, S. (1990) "Sampling of Atmospheric Carbonyls with Small DNPH-coated C_{18} Cartridges and Liquid Chromatography Analysis with Diode Array Detection," International Journal of Environmental Analytical Chemistry, 38:495-512.

Fishbein, L. (1979) Potential Industrial Carcinogens and Mutagens, Elsevier Scientific Publishing Company, New York, pp 149-150.

Highsmith, V. and Zweidinger, R. (1988) "Characterization of Indoor and Outdoor Air Associated with Residences Using Woodstoves: A Pilot Study," Environmental International, 14:213-219.

Hodgson, A. and Wooley, J. (1991) Assessment of Indoor Concentrations, Indoor Sources and Source Emissions of Selected Volatile Organic Compounds. Final Report to ARB Under Contract No. A933-063, pp 63-69.

International Commission on Radiological Protection (1975) Report of the Task Group on Reference Man, Pergamon Press, New York, p 346.

Krzymien, M. (1989) "GC-MS Analysis of Organic Vapors Emitted from Polyurethane Foam Insulation," Intern. J. of Environ. Anal. Chem., 36:193-207.

Lipari, F., Dasch, J., Scruggs, W. (1984) "Aldehyde Emissions from Woodburning Fireplaces," Environ. Sci. Technol., 18(5):326-330.

Lofroth, G., Burton, R., Forehand, L., Hammond, S., Sella, R., Zweidinger, R., Lewtas, J. (1989) "Characterization of Environmental Tobacco Smoke," Environ. Sci. Technol., 23(5):610-614.

Lund, E., Kirkland, C., Shaw, P. (1981) "Methanol, Ethanol, and Acetaldehyde Contents of Citrus Products," J. Agric. Food Chem., 29:361-363.

Metropolitan Water District of Southern California and JMM, Inc. (1989) "Disinfection By-Products in California Drinking Water," final report to California Department of Health Services.

National Research Council (1981) Formaldehyde and Other Aldehydes, National Academy Press, Washington, D.C.

Nelson, W., Bond, A., Hartlage, T., Manale, A., Wallace, L. (1988) "Preliminary Results of the Baltimore TEAM Study: I. Goals and Study Design," Proceedings of the 1988 EPA/APCA International Symposium on Measurement of Toxic and Related Air Pollutants, pp 137-142.

Nelson, W. (1990) Environmental Protection Agency, Environmental Monitoring Systems Laboratory, personal communication.

Quraishi, T. and J. Todd (1987) "Effects of Domestic Wood-Burning on Indoor Air Quality - An Overview," Indoor Air '87: Proceedings of the Fourth International Conference on Indoor Air Quality and Climate, Vol. 1; pp 488-492.

Rogozen, M., Maldonado, G., Grosjean, D., Shoshet, A., Rappaport, R. (1984) "Formaldehyde: A Survey of Airborne Concentrations and Sources," final report to ARB, contract no. A2-059-32.

Rogozen, M., Rich, H., Guttman, M., Grosjean, D. (1987) "Evaluation of Potential Toxic Air Contaminants: Phase I," final report to ARB, contract no. A4-131-32.

Sexton, K., Webber L., Hayward, S., Sextro, R. (1986) "Characterization of Particle Composition, Organic Vapor Constituents, and Mutagenicity of Indoor Air Pollutant Emissions," Environment International, 12:351-362.

Shikiya, D., Chung, S., Kahn, M., Juarros, J., Barcikowski, W. (1989) "In-Vehicle Air Toxics Characterization Study in the South Coast Air Basin," South Coast Air Quality Management District.

Tejada, S. (1990) Environmental Protection Agency, Atmospheric Sciences Research Laboratory, personal communication.

United States Department of Health and Human Services (1989) Reducing the Health Consequences of Smoking: 25 Years of Progress: A Report of the Surgeon General.

Wiley, J., Robinson, J., Piazza, T., Garrett, K., Cirksena, K., Cheng, Y., Martin, G. (1989) "Activity Patterns of California Residents," final report to ARB, contract no. A6-177-33.

World Health Organization (1987) IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans: Vol. 36, Allyl Compounds, Aldehydes, Epoxides and Peroxides, International Agency for Research on Cancer, pp 101-132.

World Health Organization (1987) "Formaldehyde," Air Quality Guidelines for Europe, WHO Regional Publications, European Series no. 23, pp 91-104.

Zweidinger, R., Dropkin, D., Stump, F., Tejada, S. (1987) "Volatile Organic Hydrocarbon and Aldehyde Composition in Raleigh, N.C. during the 1985 Woodsmoke Study," Proceedings of the 1987 EPA/APCA Symposium on the Measurements of Toxic and Related Air Pollutants, pp 579-584.

Zweidinger, R., Tejada, S., Highsmith, R., Westburg, H., Gage, L. (1988) "Distribution of Volatile Organic Hydrocarbons and Aldehydes During the IACP Boise, Idaho residential Study," Proceedings of the 1988 EPA/APCAS International Symposium on the Measurement of Toxic and Related Air Pollutants, pp 814-820.

Zweidinger, R. (1990) Environmental Protection Agency, Atmospheric Sciences Research Laboratory, personal communication.

APPENDIX F

INFORMATION REQUEST LETTER WITH ATTACHMENTS AND RESPONSES

STATE OF CALIFORNIA
George Deukmejian, Governor

AIR RESOURCES BOARD
1102 Q STREET
P.O. BOX 2815
SACRAMENTO, CA 95812

March 24, 1989

Dear Sir or Madam:

Request for Information Regarding Acetaldehyde

I am writing to request information on the atmospheric chemistry, sources, total exposure to, and health effects of acetaldehyde. The California Air Resources Board (ARB) is requesting this information as part of our toxic air contaminant program. This program is mandated by California Health and Safety Code Section 39650 eq seq. A summary of this program is contained in Attachment 1, which also describes the statutory basis for the program.

The information that you provide will be considered in an evaluation of acetaldehyde as a candidate toxic air contaminant; the evaluation will be conducted jointly by the ARB and the state's Department of Health Services (DHS). As part of the evaluation, we will consider all available health and exposure information regarding acetaldehyde.

In February 1989, we conducted a reference search on acetaldehyde exposure and health effects using several data bases from the Dialogue data retrieval system. These references include material published from 1972 to late 1988. The attached bibliography (Attachment 2) lists the most recent and relevant references from this information search. We are requesting additional, pertinent information on acetaldehyde health effects and exposure, including any material that may not be available to the public or that is not included in the attached bibliography.

We are also requesting information relevant to human exposure to acetaldehyde in California through media other than outdoor ambient air. Specifically, we are requesting information concerning total exposure to acetaldehyde through inhalation, ingestion, and skin absorption, with special emphasis on inhalation exposure in the indoor air environment. This includes acetaldehyde concentrations in various media, consumption information, bioabsorption rates, body burden studies, etc. We are also requesting data on emission sources responsible for acetaldehyde exposures in media other than outside air.

The information that you provide with the exception of trade secrets may be released to the public. The Air Resources Board's procedure for handling information claimed to be trade secrets is explained in Attachment 3.

If you believe that any of the information you are providing is a trade secret or otherwise exempt from disclosure under any other provision of law, you should identify it as such at the time of submission (Health and Safety Code Section 39660(e)). The ARB may ask you to provide documentation of your claim of trade secret or exemption at a later date.

I would appreciate receiving by April 28, 1989 any relevant information you wish to submit. Your help in expediting our review will be greatly appreciated. Please send the information to the attention of:

Robert Barham, Chief
Toxic Air Contaminant Identification Branch
Re: Acetaldehyde
California Air Resources Board
P. O. Box 2815
Sacramento, CA 95812

If you have further questions regarding indoor acetaldehyde exposure, please contact John Batchelder at (916) 323-1505. For other questions, please contact Robert Rood at (916) 445-6138.

If you are not the person to whom this request should be addressed, please forward it to the appropriate person in your organization. Also, please let us know whether you would like to continue to receive information inquiries for other candidate substances, and if not, if there is anyone in your organization to whom such requests should be sent.

Sincerely,

Peter D. Venturini, Chief
Stationary Source Division

Attachments

cc: Michael Lipsett, Department of Health Services
Rex Magee, Department of Food and Agriculture
Roxanne Jacques, President, California Air Pollution Control Officers Association
Stew Wilson, Executive Secretary, California Air Pollution Control Officers Association
David Howekamp, EPA, Region 9
Mike Stenburg, EPA, Region 9
Kathy Diehl, EPA, Region 9
Assemblywoman Sally Tanner, Chairwoman Committee on Toxic Materials
Senator Ralph Dills, Chairman Committee on Governmental Organization
Senator Art Torres, Chairman, Committee on Toxics and Public Safety Management
Robert D. Barham, Air Resources Board
John Batchelder, Air Resources Board
Robert Rood, Air Resources Board
Scientific Review Panel Members
Air Pollution Control Officers

Attachment 1
February 1989

State of California
AIR RESOURCES BOARD

Toxic Air Contaminant: Program Statutory Basis and Process

California's toxic air contaminant program is established by Health and Safety Code sections 39650-39674. This law sets forth the process for:

1. assessing the risk posed by substances;
2. identifying by regulation substances determined to be toxic air contaminants; and
3. managing the risk by adopting control measures for the identified toxic air contaminant

The law defines a toxic air contaminant as an air pollutant which may cause or contribute to an increase in mortality or an increase in serious illness, or which may pose a present or potential hazard to human health.

Two reports are prepared during this risk assessment and risk management process. The first report, which is prepared by the Air Resources Board (ARB) staff and the Department of Health Services (DHS), contains information on the health effects of the substance. This report is used to determine whether a substance should be identified as a toxic air contaminant. We refer to this report as the "ID" report. The second report, which is prepared by the ARB staff after a substance is identified as a toxic air contaminant, is on the need for and appropriate degree of regulation of that substance . We refer to this report as the "regulatory needs" report. Both reports are made available to the public for review and comment.

Before the ARB can formally identify a substance as a toxic air contaminant, several steps must be taken. First, the ARB must request the DHS to evaluate the health effects of the candidate substance. Second, the ARB staff must prepare the "ID" report that includes the estimate of exposure levels and also the health effects evaluation and then submit the report to a Scientific Review Panel for its review. The report submitted to the Panel will be made available to the public. Information submitted by interested parties will be considered in the report to the Panel. The Panel reviews the sufficiency of the information, methods, and data used by the DHS in its evaluation. Last, after review by the Scientific Review Panel, the report with the written findings of the Panel will be considered by the ARB and will be the basis for any regulatory action to identify a substance as a toxic air contaminant.

The attached information request is the first step in preparing the "ID" report on whether a substance should be identified as a toxic air contaminant. Before either the ARB or the DHS begin their analysis of a substance, the ARB provides an opportunity for interested parties to submit information on the atmospheric chemistry, sources, exposure to and health effects of that substance. You are encouraged to submit any information that you believe would be important in DHS' and ARB's evaluation.

Attachment 2
February 1989

State of California
AIR RESOURCES BOARD

Bibliography: Acetaldehyde Exposure and Health Effects

ACETALDEHYDE ATMOSPHERIC EXPOSURE RELATED REFERENCES - 2/22/89

Ahrenholz, S. H.
Health Hazard Evaluation Report HETA 87-266-1860, U.S. Department of the Treasury, Bureau of Engraving and Printing, Washington, DC.
National Inst. for Occupational Safety and Health, Cincinnati, OH. Hazard Evaluations and Technical Assistance Branch.
Corp. Source Codes: 052678009
Report No.: HETA-87-266-1860
Jan 88 45p
Journal Announcement: GRAI8820

Ahrenholz, S. H.; Lipscomb, J.
Health Hazard Evaluation Report No. HETA 82-096-1259, Kaiser Aluminum and Chemical Corporation, Ravenswood, West Virginia.
National Inst. for Occupational Safety and Health, Cincinnati, OH.
Corp. Source Codes: 052678000
Report No.: HETA-82-096-1259
Feb 83 32p
Journal Announcement: GRAI8419

Anderson, L. G.; Machovec, C. M.; Lanning, J. A.
Effects of the Mandated High Oxygen Fuels Program on the Concentrations of Formaldehyde and Acetaldehyde in Denvers Air.
Univ Colorado, Dpt Chem/Denver//CO/80204
Abstracts of Papers of the American Chemical Society, 1988, v196, Sep, p 40-ENVR

Anonymous
American Chemical Society, Division of Environmental Chemistry, 191st National Meeting, Volume 26, Number 1.
Source: National Meeting - American Chemical Society, Division of Environmental Chemistry 191st v 26 n 1. Publ by ACS, Washington, DC, USA 292p
Publication Year: 1986

Apol, A.
Health Hazard Evaluation Report No. HETA 82-137-1264, Regional Transportation District, Denver, Colorado.
National Inst. for Occupational Safety and Health, Cincinnati, OH.
Corp. Source Codes: 052678000
Report No.: HETA-82-137-1264
Feb 83 27p
Journal Announcement: GRAI8419

Atkinson, G. H.
Photoacoustic Detection of NO2, H2CO, and CH3CHO.
(Final rept.)
Arizona Univ., Tucson.
Corp. Source Codes: 000951000
Sponsor: Coordinating Research Council, Inc., Atlanta, GA.
Report No.: CRC-APRAC-CAPA-17-19-80-04
1986 65p
Sponsored by Coordinating Research Council, Inc., Atlanta, GA.
Journal Announcement: GRAI8714

Atkinson, George H.
Detection of Atmospheric Aldehydes by Photoacoustic Laser Spectroscopy.
(Annual rept.)
Syracuse Univ., NY. Dept. of Chemistry.
Corp. Source Codes: 013939022
Sponsor: Coordinating Research Council, Inc., Atlanta, GA.
Report No.: CRC-APRAC-CAPA-17-80
1981 23p
Journal Announcement: GRAI8304

Atkinson, R.; Lloyd, A. C.
Evaluation of Kinetic and Mechanistic Data for Modeling of Photochemical Smog.
California Univ., Riverside.
Corp. Source Codes: 005433000;
Sponsor: Environmental Research and Technology, Inc., Newbury Park, CA.
1984 130p
Prepared in cooperation with Environmental Research and Technology, Inc., Newbury Park, CA.
Included in Jnl. of Physical and Chemical Reference Data, v13 n2 p315-444 1984.
Journal Announcement: GRAI8505

Bagnall, G. N.; Sidebottom, H. W.
Photooxidation of Acetaldehyde.
Location: Chem. Dep., Univ. Coll., Dublin, Ire.,
Journal: Comm. Eur. Communities, (Rep.) EUR Date: 1984 Number: EUR 9436, Phys.-Chem. Behav. Atmos. Pollut. Pages: 188-93

Brander, Susann M.; Johansson, Gudrun I.; Stenius, Per J.; Kronberg, Bengt G.
Reactive Foams for Air Purification.
Inst for Surface Chemistry, Sweden,
Env Science & Technology, Apr 84, v18, n4, p224 (7)

Cantrell, Christopher A.; Davidson, James A.; Busarow, Kerry L.;
Calvert, Jack G.
The Acetaldehyde-Nitrogen Trioxide Reaction with Possible Nighttime Peroxyacetylnitrate Generation.
Location: Natl. Cent. Atmospheric Res., Boulder, CO, USA
Journal: J. Geophys. Res., D: Atmos. Date: 1986 Volume: 91 Number: D5
Pages: 5347-53

Chien, J. C.; Lu, P. H.
Radiolysis of Poly(chloroacetaldehyde), A Positive E-Beam Resist. (Technical rept.)
Massachusetts Univ., Amherst. Dept. of Polymer Science and Engineering.
Corp. Source Codes: 010574072; 406718
Report No.: TR-5
Jun 87 36p
Journal Announcement: GRAI8722

Chiorboli, C.; Maldotti, A.; Bignozzi, C. A.; Carassiti, V.
Atmospheric Photochemistry: Kinetics and Mechanism of Reactions Between Aromatic Olefins and Hydroxyl Radical.
CNR, Centro di Fotochimica, Ferrara, Italy
Source: Commission of the European Communities, (Report) EUR 7624. Publ by D. Reidel Publ Co, Dordrecht, Neth and Boston, Mass, USA p 228-233
Publication Year: 1982

Chrostek, Walter J.; Shoemaker, William E.
Health Hazard Evaluation Report No. HETA-81-298-944, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania.
National Inst. for Occupational Safety and Health, Cincinnati, OH.
Corp. Source Codes: 052678000
Report No.: HETA-81-298-944
Aug 81 10p
Journal Announcement: GRAI8305

Claxton, L. D.; Cupitt, L. T.; Kleindienst, T. E.; Shepson, P. B.;
Edney, E. O.
Mutagenic Activity of the Products of Propylene Photooxidation. (Journal article)
Health Effects Research Lab., Research Triangle Park, NC.
Corp. Source Codes: 048097000;
Sponsor: Northrop Services, Inc./Environmental Sciences, Research Triangle Park, NC.
Report No.: EPA/600/J-85/516
c1985 10p
Journal Announcement: GRAI8810

Davidson, James A.; Cantrell, Christopher A.; Shetter, Richard E.; Calvert, Jack G.
NO//3-Radical Reactions with Some Inorganic Species.
Natl Cent for Atmospheric Research, Boulder, CO, USA
Source: National Meeting - American Chemical Society, Division of Environmental Chemistry 189th, v 25 n 1. Publ by ACS, Washington, DC, USA p 149-150
Publication Year: 1985

De Andrade, Jailson B.; Miguel, Antonio H.
Determination of Carbonyl Compounds in Exhaust Gases from Alcohol-Fueled Vehicles Equipped with Three-Way Catalytic Converters.
Location: Dep. Quim., Pontif. Univ. Catol. Rio de Janeiro, Rio de Janeiro, Brazil
Journal: Int. J. Environ. Anal. Chem. Date: 1985 Volume: 21 Number: 3 Pages: 229-37

Dilling, Wendell L.; Bredeweg, Corwin J.; Tefertiller, Nancy B.
Organic Photochemistry: Simulated Atmospheric Photodecomposition Rates of Methylene Chloride, 1,1,1-Trichloroethane, Trichloroethylene, Tetrachloroethylene, and Other Compounds.
Dow Chemical Co.
Env Science & Technology, Apr 76, v10, n4, p351 (6)

Dodge, M. C.
Combined Effects of Organic Reactivity and NMHC/NOx Ratio on Photochemical Oxidant Formation - A Modeling Study.
(Journal article)
Environmental Sciences Research Lab., Research Triangle Park, NC.
Corp. Source Codes: 046904000
Report No.: EPA/600/J-84/403
1984 11p
Pub. in Atmospheric Environment, v18 n8 p1657-1665 1984.
Journal Announcement: GRAI8715

Edney, Edward; Mitchell, Steven; Bufalini, Joseph J.
Atmospheric Chemistry of Several Toxic Compounds.
(Final rept.)
Environmental Sciences Research Lab., Research Triangle Park, NC.
Corp. Source Codes: 046904000
Report No.: EPA-600/3-82-092
Nov 82 120p
Journal Announcement: GRAI8307

Fung, Kochy K.; Wright, Barbara J.
Measurements of Formaldehyde, Acetaldehyde and Peroxyacetyl Nitrate for the SCCCAMP.
Location: ERT, Inc., Newbury Park, CA, USA
Journal: Proc. - APCA Annu. Meet. Date: 1987 Volume: 80th Number: Vol. 4 Pages: 87/67.1, 16 pp.

Gaffney, J. S.; Hall, J. H.; Rundberg, R. S.; Erdal, B. R.
Peroxyacyl Nitrates: Their Physical and Chemical Properties.
Los Alamos Natl Lab, Los Alamos, NM, USA
Source: National Meeting - American Chemical Society, Division of Environmental Chemistry v 26 n 2 1986. Publ by ACS, Washington, DC, USA p 165-166
Publication Year: 1986

Gesser, Hyman D.
Abatement of Indoor Formaldehyde Vapor and other Indoor Gaseous Pollutants.
Location: Can.
Patent: Canada; CA 1241524 A1 Date: 1988
Pages: 7 pp.

Gold, A.; Dube, C. E.; Perni, R. B.
Solid Sorbent Sampling Acrolein in Air.
Harvard School of Public Health.
Analytical Chemistry, Nov 78, v50, n13, p1839(3)

Grosjean, Daniel
Aldehydes, Carboxylic Acids and Inorganic Nitrate During NSMCS.
Daniel Grosjean & Associates Inc, Ventura, CA, USA
Source: Atmospheric Environment v 22 n 8 1988 p 1637-1648
Publication Year: 1988

Grosjean, Daniel
Atmospheric Reactions of Styrenes and Peroxybenzoyl Nitrate.
Daniel Grosjean & Associates Inc, Camarillo, CA, USA
Source: Science of the Total Environment v 46 Nov 1985 p 41-59
Publication Year: 1985

Grosjean, Daniel
Atmospheric Reactions of Ortho Cresol: Gas Phase and Aerosol Products.
Environmental Research & Technology Inc, Westlake Village, Calif, USA
Source: Atmospheric Environment v 18 n 8 1984 p 1641-1652
Publication Year: 1984

Grosjean, Daniel
Formaldehyde and Other Carbonyls in Los Angeles Ambient Air.
Env Research & Technology, CA,
Env Science & Technology, May 82, v16, n5, p254 (9)

Grosjean, Daniel
Measurements of Inorganic Nitrate, Aldehydes and Carboxylic Acids During the NSMCS Study.
Daniel Grosjean & Associates Inc, Camarillo, CA, USA
Source: National Meeting - American Chemical Society, Division of Environmental Chemistry v 26 n 2 1986. Publ by ACS, Washington, DC, USA p 361-363
Publication Year: 1986

Hendry, D. G.; Baldwin, A. C.; Golden, D. M.
Computer Modeling of Simulated Photochemical Smog.
SRI INTL, CA,
NTIS Report PB80-169675, FEB 80 (222)

Herrick, Robert F.
Industrial Hygiene Survey Report on Urea Formaldehyde Foam Insulation Manufacturing.
(Industrywide study)
National Inst. for Occupational Safety and Health, Cincinnati, OH. Div. of Surveillance, Hazard Evaluations and Field Studies.
Corp. Source Codes: 052678007
Report No.: IWS-69.16
8 Jul 80 25p
Journal Announcement: GRAI8301

Hesselbrock, Victor M.; Shaskan, Edward G.
Endogenous Breath Acetaldehyde Levels Among Alcoholic and Non-Alcoholic Probands: Effect of Alcohol Use and Smoking.
Location: Health Cent., Univ. Connecticut, Farmington, CT, 06032, USA
Journal: Prog. Neuro-Psychopharmacol. Biol. Psychiatry Date: 1985
Volume: 9 Number: 3 Pages: 259-65

Hoshika, Y.
Simple and Rapid Gas-Liquid-Solid Chromatographic Analysis of Trace Concentrations of Acetaldehyde in Urban Air.
Journal of Chromatography, 1977 Vol. 137, No. 2 (July 21), p. 455

Huang, S. S.
Application of Vitreous Silica Capillary Columns for GC/MS Analysis of Aldehydes in Air as Oxazolidine Derivatives.
US Bur of Engraving & Printing, Technical Services Div, Washington, DC, USA
Source: National Meeting - American Chemical Society, Division of Environmental Chemistry 186th, v 23 n 2 1983. Publ by ACS, Washington, DC, USA p506-508
Publication Year: 1983

Ibusuki, Takashi; Takeuchi, Koji
Influences of Particulate Matters in the Atmosphere on Photochemical Reactions among Gaseous Pollutants.
Natl Research Inst for Pollution & Resources, Jpn
Source: Kogai Shigen Kenkyusho Hokoku/Report of the National Research Institute for Pollution and Resources n 31 Mar 1984 64p
Publication Year: 1984

Ito, Kenichi; Kurata, Kouichi
Photochemical Reaction of the Exhaust Gas From an Alcohol-Fueled Engine (Formation of Aldehydes).
Hokkaido Univ, Dep of Mechanical Engineering, Sapporo, Jpn
Source: Bulletin of the JSME v 28 n 243 Sep 1985 p 2028-2033
Publication Year: 1985

Ito, Kenichi; Yano, Toshiaki; Takahata, Toshio
Formaldehyde and Acetaldehyde in Exhaust Gases Emitted from an Ethanol-Fueled S. I. Engine.
Location: Dep. Mech. Eng., Hokkaido Univ., Sapporo, Japan,
Journal: Bull. JSME Date: 1984 Volume: 27 Number: 227 Pages: 945-51

Jauhonen, P.; Baraona, E.; Miyakawa, H.; Lieber, C. S.
Origin of Breath Acetaldehyde During Ethanol Oxidation. Effect of Long-Term Cigarette Smoking.
J Lab Clin Med Dec 1982, 100 (6) p908-16, ISSN 0022-2143
Journal Code: IVR

Klus, H.; Begutter, H.; Nowak, A.; Pinterits, G.; Ultsch, I.; Wihlidal, H.
Indoor Air Pollution Due to Tobacco Smoke Under Real Conditions. Preliminary results.
Tokai J Exp Clin Med Aug 1985, 10 (4) p331-40, ISSN 0385-0005
Journal Code: VZM

Kyle, B. G.; Eckhoff, N. D.
Odor Removal from Air by Adsorption on Charcoal.
Kansas State Univ.
NTIS REPORT PB-236 928, SEP 74 (109)

Lipari, Frank; Dasch, Jean M.; Scruggs, William F.
Aldehyde Emissions From Wood-Burning Fireplaces.
General Motors Research Labs, MI.
Env Science & Technology, May 84, v18, n5, p326 (5)

Lunsford, R. A.; Okenfuss, J. R.; Shulman, S. A.
Diesel Exhaust/Coal Dust Exposure Study: Characterization of Selected Vapor-Phase Organic Emissions.
National Inst. for Occupational Safety and Health, Cincinnati, OH.
Corp. Source Codes: 052678000
28 Feb 84 36p
Journal Announcement: GRAI8713

Maeda, Yasuaki; Kada, Norio; Suetaka, Tohru; Munemori, Makoto
Formation of Aldehydes in Photochemical Smog Reaction of Hydrocarbon-NO//x.
Univ of Osaka Prefecture, Sakai, Jpn
Source: Chemistry Express v 3 n 5 May 1988 p 259-262
Publication Year: 1988

Meng, Z.; Tanner, R. L.
Rapid Determination of PpB Levels of Aldehydes in Ambient Air by High-Performance Liquid Chromatography.
Brookhaven National Lab., Upton, NY.
Corp. Source Codes: 004545000; 0936000
Sponsor: Department of Energy, Washington, DC.
Report No.: BNL-51725
Jun 83 60p
Journal Announcement: GRAI8420; NSA0900

Moortgat, G. K.; Burrows, J. P.; Schneider, W.; Tyndall, G. S.; Cox, R. A.
A Study of the Hydroperoxo + Acetaldehyde Reaction in the Photolysis of Acetaldehyde, and its Consequences for Atmospheric Chemistry.
Location: Air Chem. Div., Max-Planck-Inst. Chem., D-6500, Mainz, Fed. Rep. Ger.
Journal: Comm. Eur. Communities, (Rep.) EUR Date: 1987 Number: EUR 10832, Phys.-Chem. Behav. Atmos. Pollut. Pages: 271-81

Moortgat, Geert K.; McQuigg, Robert D.
A FTIR Spectroscopic Study of the Photooxidation of Acetaldehyde in Air.
Location: Air Chem. Div., Max-Planck-Inst. Chem., D6500, Mainz, Fed. Rep. Ger.
Journal: Comm. Eur. Communities, (Rep.) Eur Date: 1984 Number: EUR 9436, Phys.-Chem. Behav. Atmos. Pollut. Pages: 194-204

Motohashi, Chuichi
Removal of Lower Aldehydes Using Adsorbents Containing Thio Groups.
Location: Japan,
Assignee: Sumitomo Chemical Co., Ltd.
Patent: Japan Kokai Tokkyo Koho; JP 88200819 A2; JP 63200819
Date: 1988 Pages: 3pp

Mueller, R.; Schurath, U.
Entwicklung eines Geraetes zur kontinuierlichen Messung der Photodissoziations-Geschwindigkeit von Aldehyden in der Atmosphaere durch Nachweis des erzeugten CO. Abschlussbericht. (Development of a Continuous Actinometer for Aldehydes in Air based on Detection of the Photochemically Produced Carbon Monoxide. Final Report.)
Gesellschaft fuer Strahlen- und Umweltforschung m.b.H. Muenchen, Neuherberg (Germany, F.R.).
Corp. Source Codes: 078026000
Report No.: GSF-BPT--11/86
1986 65p
In German, With 23 figs., 2 tabs., 23 refs.
Country of Publication: Germany, Federal Republic of
Journal Announcement: GRAI8723

National Inst. for Occupational Safety and Health, Cincinnati, OH.
NIOSH (National Institute for Occupational Safety and Health) Manual of Analytical Methods (Third Edition). Second Supplement.
Corp. Source Codes: 052678000
Report No.: DHEW/PUB/NIOSH-87-117
15 Aug 87 286p
Journal Announcement: GRAI8816

Niki, H.; Maker, P. D.; Savage, C. M.; Breitenbach, L. P.
FTIR Study of the Kinetics and Mechanism for Chlorine-Atom-Initiated Reactions of Acetaldehyde.
Location: Res. Staff, Ford Mot. Co., Dearborn, MI, 48121, USA
Journal: J. Phys. Chem. Date: 1985 Volume: 89 Number: 4 Pages: 588-91

Olson, Keith L.; Swarin, Stephen J.
Determination of Aldehydes and Ketones by Derivatization and Liquid Chromatography-Mass Spectrometry.
Location: Anal. Chem. Dep., Gen. Mot. Res. Lab., Warren, MI, 48090-9055, USA
Journal: J. Chromatogr. Date: 1985 Volume: 333 Number: 2 Pages: 337-47

Olson, Terese M.; Torry, Lori A.; Hoffmann, Michael R.
Kinetics of the Formation of Hydroxyacetaldehyde-Sulfur(IV) Adducts at Low pH.
Location: Environ. Eng. Sci. W. M. Keck Lab., California Inst. Technol., Pasadena, CA, 91125, USA
Journal: Environ. Sci. Technol. Date 1988 Volume: 22 Number: 11 Pages: 1284-9

Paolo, M.; Renzo, B.
Determination of Metaldehyde in Workroom Air.
Bull Environ Contam Toxicol Apr 1983, 30 (4) p479-84, ISSN 0007-4861
Journal Code: BFN

Peltonen, K.; Pfaffli, P.; Itkonen, A.
Determination of Aldehydes in Air as Dimethone Derivatives by Gas Chromatography with Electron-Capture Detection.
J Chromatogr Dec 19 1984, 315 p412-6, ISSN 0021-9673
Journal Code: HQF

Pervier, J. W.; Barley, R. C.; Field, D. E.; Friedman, B. M.; Morris, R. B.
Survey Reports on Atmospheric Emissions from the Petrochemical Industry, Air Products & Chemicals, PA.
NTIS Reports PB-244 957-8, Mar-Apr 74 (593)

Rocklin, Roy D.
Ion Chromatography with Pulsed Amperometric Detection. Simultaneous Determination of Formic Acid, Formaldehyde, Acetaldehyde, Propionaldehyde, and Butyraldehyde.
Location: Dionex Corp., Sunnyvale, CA, 94088-3603, USA
Journal: Adv. Chem. Ser. Date: 1985 Volume: 210 Number: Formaldehyde
Pages: 13-21

Rogers, Jerry D.
Infrared Intensities of Acetaldehyde Fundamental Bands.
Location: Environ. Sci. Dep., Gen. Mot. Res. Lab., Warren, MI, 48090, USA
Journal: J. Quant. Spectrosc. Radiat. Transfer Date: 1985 Volume: 34
Number: 1 Pages: 27-32

Rogozen, M. B.; Rich, H. E.; Guttman, M. A.; Grosjean, D.
Evaluation of Potential Toxic Air Contaminants. Phase 1. Final Report. (Rept. for Jul 85-Sep 86)
Science Applications International Corp., Manhattan Beach, CA.
Corp. Source Codes: 091348000;
Sponsor: Grosjean (Daniel) and Associates, Ventura, CA.; California State Air Resources Board, Sacramento.
Report No.: SAIC-76/1035; ARB-R-88/333
23 Dec 87 582p
Journal Announcement: GRAI8813

Rudling, L.; Ahling, B.; Loefroth, G.
Chemical and Biological Characterization of Emissions from Combustion of Wood-Chips in a Small Central Heating Furnace and from Combustion of Wood in Closed Fireplace Stoves.
Corp. Source Codes: 9999999
Report No.: SNV-PM-1331
Sep 80 79p
In Swedish.
Journal Announcement: GRAI8221; NSA0000

Rudling, L.; Loefroth, G.
Emissions from Combustion of Peat and Wood-Chips.
National Swedish Environment Protection Board, Solna.
Corp. Source Codes: 043323000; 9698090
Report No.: SNV-PM-1449
1981 96p
In Swedish.
Journal Announcement: GRAI8317; NSA0800

Salas, Louis J.; Singh, Hanwant B.
Measurements of Formaldehyde and Acetaldehyde in the Urban Ambient Air.
SRI-Int, Menlo Park, CA, USA
Source: Atmospheric Environment v 20 n 6 1986 p 1301-1304
Publication Year: 1986

Schubert, B.; Schmidt, U.; Ehhalt, D. H.
Sampling and Analysis of Acetaldehyde in Tropospheric Air.
Location: Inst. Chem., Kernforschungsanlage Juelich G.m.b.H., D-5170, Juelich, Fed. Rep. Ger.
Journal: Comm. Eur. Communities, (Rep.) EUR Date: 1984 Number: EUR 9436, Phys.-Chem. Behav. Atmos. Pollut. Pages: 44-52

Schulam, P.; Newbold, R.; Hull, L. A.
Urban and Rural Ambient Air Aldehyde Levels in Schenectady, New York and on Whiteface Mountain, New York.
Location: Dep. Chem., Union Coll., Schenectady, NY, 12308, USA
Journal: Atmos. Environ. Date: 1985 Volume: 19 Number: 4 Pages: 623-6

Schutte, Norbert P.
Hazard Evaluation and Technical Assistance Report No. TA-77-11, Xomed Company, Cincinnati, Ohio.
National Inst. for Occupational Safety and Health, Cincinnati, OH. Hazard Evaluations and Technical Assistance Branch.
Corp. Source Codes: 052678009
Report No.: TA-77-11
Nov 77 18p
Journal Announcement: GRAI8219

Serth, R. W.; Tierney, D. R.; Hughes, T. W.
Source Assessment: Acrylic Acid Manufacture State of the Art.
Monsanto Research Corp, Ohio.
EPA Report EPA-600/2-78-004W Aug 78, (69)

Shaskan, Edward G.; Dolinsky, Zelig S.
Elevated Endogenous Breath Acetaldehyde Levels among Abusers of Alcohol and Cigarets.
Location: Health Cent., Univ. Connecticut, Farmington, CT, 06032, USA
Journal: Prog. Neuro-Psychopharmacol. Biol. Psychiatry Date: 1985 Volume: 9 Number: 3 Pages: 267-72

Shepson, P. B.; Kleindienst, T. E.; Edney, E. O.
Production of Mutagenic Compounds as a Result of Urban Photochemistry.
(Interim rept. Jul 83-Feb 87)
Northrop Services, Inc., Research Triangle Park, NC.
Corp. Source Codes: 058582000
Sponsor: Environmental Protection Agency, Research Triangle Park, NC.
Atmospheric Sciences Research Lab.
Report No.: EPA/600/3-87/020
Jun 87 86p
Sponsored by Environmental Protection Agency, Research Triangle Park, NC.
Atmospheric Sciences Research Lab.
Journal Announcement: GRAI8718

Silveira de Queiroz, Rogerio; De Faro Orlando, Alcir;
Bittencourt, Aloysio M. Bomtempo
Pollutant Emissions from Internal Combustion Engines using Ethyl
Alcohol-Water Blends and Gasoline as Fuels.
Location: Fed. Univ. Espirito Santo, Brazil
Journal: Altern. Energy Sources, (Proc. Miami Int. Conf.), 6th Editor:
Veziroglu, T. Nejat (Ed), Date: 1985 Volume: 4, Pages: 201-15

Singh, Hanwant B.; Hanst, Philip L.
Peroxyacetyl Nitrate (PAN) in the Unpolluted Atmosphere: An Important
Reservoir for Nitrogen Oxides.
(Journal article)
SRI International, Menlo Park, CA.
Corp. Source Codes: 055876000
Report No.: EPA-600/J-81-628
c1981 6p
Pub. in Geophysical Research Letters, v8 n8 p941-944 1981.
Journal Announcement: GRAI8321

Sivakumar, N.; Hall, G. E.; Houston, P. L.; Hepburn, J. W.; Burak, I.
State-Resolved Photodissociation of OCS Monomers and Clusters.
Cornell Univ, Dept Chem/Ithica//NY/14853; Univ Waterloo, Ctr Molec Beams
& Laser Chem/Waterloo N2L 3G1/Ontario/Canada/; TEL AVIV Univ,Sackler
Fac Sci, Sch Chem/IL-69978TEL AVIV//Israel/
Journal of Chemical Physics, 1988, v88, n6, p3692-3708

Springer, Charles; Thibodeaux, Louis J.; Hedden, Terry
Thermal Effects on the Emission of Volatile Organic Chemicals from
Surface Waste-Water Impoundments in the Absence of Wind: Laboratory
Measurements.
Location: Univ. Arkansas, Fayetteville, AR, USA
Journal: Proc. - APCA Annu. Meet. Date: 1984 Volume: 77th Number: Vol. 1
Pages: 84-3.4, 22 pp.

Suyama, Yoshiaki; Nakazawa, Makoto; Honda, Hisao; Makino, Hiroshi; Saiki, Yoshio; Kajino, Tadashi; Sano, Sadao
Study on Pollutant Emission from Methanol fueled Vehicles (Part 2); Emission Characteristics of Formaldehyde and the Like.
Location: Spec. Air Qual. Dep., Kanagawaken Kogai Cent., Yokohama, Japan, Journal: Kenkyu Hokoku - Kanagawa-ken Kogai Senta Date: 1988 Volume: 10, Pages: 34-41 Language: Japanese

Thibodeaux, Louis J.; Parker, David G.; Heck, Howell H.
Measurement of Volatile Chemical Emissions from Wastewater Basins.
(Final rept.)
Arkansas Univ., Fayetteville.
Corp. Source Codes: 000952000
Sponsor: Industrial Environmental Research Lab.-Cincinnati, OH.
Report No.: EPA-600/2-82-095
Nov 82 60p
Journal Announcement: GRAI8306

Wesely, M. L.
Improved Parameterizations for Surface Resistance to Gaseous Dry Deposition in Regional-Scale, Numerical Models.
(Final rept.)
Argonne National Lab., IL. Biological and Medical Research Div.
Corp. Source Codes: 001960006
Sponsor: Environmental Protection Agency, Research Triangle Park, NC.
Atmospheric Sciences Research Lab.; Department of Energy, Washington, DC.
Report No.: EPA/600/3-88/025
Jun 88 59p
Journal Announcement: GRAI8820

Whitten, G. Z.; Killus, J. P.; Johnson, R. G.
Modeling of Simulated Photochemical Smog with Kinetic Mechanisms.
(Final rept. Sep 79-Aug 82)
Systems Applications, Inc., San Rafael, CA.
Corp. Source Codes: 063315000
Sponsor: Environmental Sciences Research Lab., Research Triangle Park, NC.
Report No.: EPA-600/3-83-043
May 83 329p
Journal Announcement: GRAI8323

Williams, R. G.
Determination of Chloroacetaldehyde in Air by Differential Pulse Polarography.
Anal Chem Oct 1982, 54 (12) p2121-2, ISSN 0003-2700 Journal Code: 4NR

Yano, Toshiaki; Ito, Kenichi; Takahata, Toshio
Formaldehyde and Acetaldehyde in Exhaust Gases Emitted from an Ethanol Fueled S.I. Engine. Clarification of the Formation Mechanism by Reactor Experiments.
Location: Dep Tech. Educ., Kagoshima Univ., Kagoshima, Japan, 890
Journal: Bull. JSME Date: 1986 Volume: 29 Number: 255 Pages: 3028-35

Zey, J. N.; Todd, W. F.
Health Hazard Evaluation Report HETA 85-308-1829, Kardon Industries, St. Paris, Ohio.
National Inst. for Occupational Safety and Health, Cincinnati, OH. Hazard Evaluations and Technical Assistance Branch.
Corp. Source Codes: 052678009
Report No.: HETA-85-308-1829
Sep 87 21p
Journal Announcement: GRAI8810

ACETALDEHYDE HEALTH RELATED REFERENCES - 2/22/89

Ahrenholz, S. H.; Lipscomb, J.
Health Hazard Evaluation Report No. HETA 82-096-1259, Kaiser Aluminum and Chemical Corporation, Ravenswood, West Virginia.
National Inst. for Occupational Safety and Health, Cincinnati, OH.
Corp. Source Codes: 052678000
Report No.: HETA-82-096-1259
Feb 83 32p
Journal Announcement: GRAI8419

Andronova, L. M.; Ushakova, M. M.; Kudriavtsev, R. V.; Barkov, N. K.
[Significance of the Sex Differences of Experimental Animals for Ethanol and Acetaldehyde Pharmacokinetics.]
Znachenie polovykh razlichii eksperimental'nykh zhivotnykh dlia farmakokinetiki etanola i atsetal'degida.
Farmakol Toksikol Sep-Oct 1982, 45 (5) p101-5, ISSN 0014-8318
Journal Code: ETR

Anonymous
American Chemical Society, Division of Environmental Chemistry, 191st National Meeting, Volume 26, Number 1.
Source: National Meeting - American Chemical Society, Division of Environmental Chemistry 191st v 26 n 1. Publ by ACS, Washington, DC, USA
292p
Publication Year: 1986

Apol, A.
Health Hazard Evaluation Report No. HETA 82-137-1264, Regional Transportation District, Denver, Colorado.
National Inst. for Occupational Safety and Health, Cincinnati, OH.
Corp. Source Codes: 052678000
Report No.: HETA-82-137-1264
Feb 83 27p
Journal Announcement: GRAI8419

Appelman, L. M.; Woutersen, R. A.; Feron, V. J.
Inhalation Toxicity of Acetaldehyde in Rats. I. Acute and Subacute Studies.
Toxicology 1982, 23 (4) p293-307, ISSN 0300-483X Journal Code: VWR

Aranyi, C.; O'Shea, W. J.; Graham, J. A.; Miller, F. J.
Effects of Inhalation of Organic Chemical Air Contaminants on Murine Lung Host Defenses.
(Journal article)
Health Effects Research Lab., Research Triangle Park, NC.
Corp. Source Codes: 048097000;
Sponsor: IIT Research Inst., Chicago, IL. Life Sciences Research Div.
Report No.: EPA/600/J-86/177
c1986 10p
Pub. in Fundamental and Applied Toxicology 6, p713-720 1986. Prepared in cooperation with IIT Research Inst., Chicago, IL. Life Sciences Research Div.
Journal Announcement: GRAI8705

Crandall, M. S.; Wilcox, T. G.
Health Hazard Evaluation Report No. HETA-87-181-0000, Graphic Packaging Corporation, Paoli, Pennsylvania.
National Inst. for Occupational Safety and Health, Cincinnati, OH. Div. of Surveillance, Hazard Evaluations and Field Studies.
Corp. Source Codes: 052678007
Report No.: HETA-87-181-0000
Jul 88 13p
Journal Announcement: GRAI8902

Dannecker, John R., Jr.; Shaskan, Edward G.; Meyer, Roger E.
Breath Acetaldehyde: Highly Sensitive Evaluation of Endogenous and Exogenous Ethanol Metabolism.
Location: Alcohol Res. Cent., Univ. Connecticut, Farmington, CT, USA
Journal: Res. Monogr. - Natl. Inst. Alcohol Abuse Alcohol. Date: 1983
Volume: 9 Number: Biol./Genet. Factors Alcohol. Pages: 49-65

Egle, John L., Jr.
Retention of Inhaled Acetaldehyde in the Dog.
Arch Envir Hlth May 1972 v24 n5 p354 (4)

Feron, V. J.; Kuper, C. F.; Spit, B. J.; Reuzel, P. G.; Woutersen, R. A.
Glass Fibers and Vapor Phase Components of Cigarette Smoke as Cofactors in Experimental Respiratory Tract Carcinogenesis.
Carcinog Compr Surv 1985, 8 p93-118, ISSN 0147-4006 Journal Code: CNU

Frostling, H.; Hoff, A.; Jacobsson, S.; Pfaffli, P.; Vainiotalo, S.; Zitting, A.
Analytical, Occupational and Toxicologic Aspects of the Degradation Products of Polypropylene Plastics.
Scand J Work Environ Health Jun 1984, 10 (3) p163-9, ISSN 0355-3140
Journal Code: UEB

Germon, J. C.; Knowles, R.
Metabolism of Acetylene and Acetaldehyde by Rhodococcus Rhodochrous.
Department of Microbiology, Macdonald College of McGill University, Ste. Anne de Bellevue, Que., Canada.
Can J Microbiol (Canada) Mar 1988, 34 (3) p242-8, ISSN 0008-4166
Journal Code: CJ3

Herrick, Robert F.
Industrial Hygiene Survey Report on Urea Formaldehyde Foam Insulation Manufacturing.
(Industrywide study)
National Inst. for Occupational Safety and Health, Cincinnati, OH. Div. of Surveillance, Hazard Evaluations and Field Studies.
Corp. Source Codes: 052678007
Report No.: IWS-69.16
8 Jul 80 25p
Journal Announcement: GRAI8301

Hesselbrock, Victor M.; Shaskan, Edward G.
Endogenous Breath Acetaldehyde Levels Among Alcoholic and Non-Alcoholic. Probands: Effect of Alcohol Use and Smoking.
Location: Health Cent., Univ. Connecticut, Farmington, CT, 06032, USA
Journal: Prog. Neuro-Psychopharmacol. Biol. Psychiatry Date: 1985
Volume: 9 Number: 3 Pages 259-65

Holford, N. H.
Clinical Pharmacokinetics of Ethanol.
Department of Pharmacology and Clinical Pharmacology, School of Medicine, University of Auckland.
Clin Pharmacokinet Nov 1987, 13 (5) p273-92, ISSN 0312-5963
Journal Code: DG5

IARC
Acetaldehyde.
IARC Monogr Eval Carcinog Risk Chem Hum Feb 1985, 36 p101-32,
Journal Code: GE4

Jauhonen, P.; Baraona, E.; Miyakawa, H.; Lieber, C. S.
Origin of Breath Acetaldehyde During Ethanol Oxidation. Effect of Long-Term Cigarette Smoking.
J Lab Clin Med Dec 1982, 100 (6) p908-16, ISSN 0022-2143
Journal Code: IVR

Jones, A. W.
Drug-Alcohol Flush Reaction and Breath Acetaldehyde Concentration: No Interference with an Infrared Breath Alcohol Analyzer.
Location: Dep. Alcohol Drug Add. Res., Karolinska Inst., S-104 01, Stockholm, Swed.
Journal: J. Anal. Toxicol. Date: 1986 Volume: 10 Number: 3
Pages: 98-101

Jones, A. W.; Skagerberg, Stefan; Borg, Stefan; Aenggaard, Erik
Time Course of Breath Acetaldehyde Concentrations During Intravenous Infusions of Ethanol in Healthy Men.
Location: St. Goerans Hosp., Karolinska Inst., Stockholm, Swed.
Journal: Drug Alcohol Depend. Date: 1984 Volume: 14 Number: 2
Pages: 113-19

Jones, A. W.; Mardh, G.; Anggard, E.
Determination of Endogenous Ethanol in Blood and Breath by Gas Chromatography-Mass Spectrometry.
Pharmacol Biochem Behav 1983, 18 Suppl 1 p267-72, ISSN 0091-3057
Journal Code: P3Q

Jones, A. W.; Neiman, J.; Hillbom, M.
Elimination Kinetics of Ethanol and Acetaldehyde in Healthy Men During the Calcium Carbimide-Alcohol Flush Reaction.
Department of Alcohol Toxicology, National Laboratory of Forensic Chemistry, University Hospital, Linkoping, Sweden.
Alcohol Alcohol 1987, Suppl 1 p213-7, ISSN 0735-0414 Journal Code: AAL

Jones, A. W.; Sato, Akio; Forsander, Olof A.
Liquid/Air Partition Coefficients of Acetaldehyde: Values and Limitations in Estimating Blood Concentrations from Analysis in Breath.
Location: Dep. Alcohol Drug Addict. Res., Karolinska Inst., 104 01, Stockholm, Swed.
Journal: Alcohol.: Clin. Exp. Res. Date: 1985 Volume: 9 Number: 5
Pages: 461-4

Kraft, J.; Kuhler, M.
Aldehyde and Ethanol Contamination from Vehicles Driven by Ethanol in Narrow Streets.
Volkswagenwerk A.G., Wolfsburg (Germany, F.R.). Abt. Forschung und Entwicklung.
Corp. Source Codes: 064598002; 9202165
Report No.: VW-FMT-8608-V/5
15 Dec 86 36p
In German.
U.S. Sales Only. Portions of this document are illegible in microfiche products.
Country of Publication: Germany, Federal Republic of
Journal Announcement: GRAI8812; NSA1300

Lam, C. W.; Casanova, M.; Heck, H. D.
Decreased Extractability of DNA from Proteins in the Rat Nasal Mucosa after Acetaldehyde Exposure.
Fundam Appl Toxicol Apr 1986, 6 (3) p541-50, ISSN 0272-0590
Journal Code: FAB

Lehmann, W. D.; Heinrich, H. C.; Leonhardt, R.; Agarwal, D. P.; Goedde, H. W.; Kneer, J.; Rating, D.
13C-Ethanol and 13C-Acetate Breath Tests in Normal and Aldehyde Dehydrogenase Deficient Individuals.
Alcohol Jul-Aug 1986, 3 (4) p227-31, ISSN 0741-8329 Journal Code: AG9

Lieber, C. S.
Metabolic Effects of Ethanol and its Interaction with Other Drugs, Hepatotoxic Agents, Vitamins, and Carcinogens - A 1988 Update.
Vet Adm Med Ctr, Liver Dis & Nutr Sect/Bronx//NY/10468; Vet Adm Med Ctr, Ctr Alcohol Res & Treatment/Bronx//NY/10468; Cuny Mt Sinai Sch Med/New York//NY/10029
Seminars in Liver Disease, 1988, v8, n1, p47-68

Lieber, C. S.
Metabolic Effects of Acetaldehyde.
Vet Adm Med Ctr,Ctr Alcoholism Res & Treatment/Bronx//NY/10468; Vet Adm Med Ctr,Liver Dis & Nutr Sect/Bronx//NY/10468; Cuny Mt Sinai Sch Med/New York//NY/10029
Biochemical Society Transactions, 1988, v16, n3, p241-247

Malik, F.; Wickramasinghe, S. N.
Haematological Abnormalities in Mice Continuously Exposed to Ethanol Vapour.
Department of Haematology, St Mary's Hospital Medical School, University of London, UK.
Br J Exp Pathol (England) Dec 1986, 67 (6) p831-8, ISSN 0007-1021
Journal Code: AWW

Messiha, F. S.; McGrath, J.
Modulation of Nitrogen Dioxide Toxicity by Lithium.
Pharmacol Biochem Behav 1984, 21 Suppl 1 p93-7, ISSN 0091-3057
Journal Code: P3Q

Neiman, Jack; Hillbom, Matti; Jones, A. W.; Benthin, Guenther;
Loewbeer, Christian; Sippel, Helmuth
Effects of a Small Dose of Ethanol and Calcium Carbimide-Induced Acetaldehyde Intoxication on human Platelet Aggregation, Associated Thromboxane Formation and Urinary Excretion of 2,3-Dinor-6-Keto Prostaglandin F1.alpha.
Location: Dep. Clin. Exp. Alcohol and Drug Addict. Res., Karolinska Inst., Stockholm, Swed.
Journal; J. Toxicol., Clin. Toxicol. Date: 1987 Volume: 25 Number 3
Pages: 185-98

Saldiva, Paulo Hilaro Nascimento; Caldeira, Marina Pires do Rio;
Massad, Eduardo; Calheiros, Debora Fernandes; Cardoso, Luiza Maria
Nunes; Bohm, Gyorgy Miklos; Saldiva, Carmen Diva
Effects of Formaldehyde and Acetaldehyde Inhalation on Rat Pulmonary Mechanics.
Location: Sch. Med., Univ. Sao Paulo, Sao Paulo, Brazil
Journal: JAT, J. Appl. Toxicol. Date: 1985 Volume: 5 Number: 5
Pages: 288-92

Savina, V. P.; Anisimov, B. V.
Changes in Functional Parameters of Animals during Long-Term Inhalation of Acetic Acid.
Joint Publications Research Service, Arlington, VA.
Corp. Source Codes: 056306000; J1957394
Sponsor: National Aeronautics and Space Administration, Washington, DC.
c1988 6p
In Its JPRS Report: Science and Technology. USSR: Space Biology and Aerospace Medicine, v21 n4 p79-84 Jul-Aug 87. Trans. into English form.
Kosmicheskaya Biologiya I Aviakosmicheskaya Meditsina (Moscow, USSR), v21
n4 p56-60 Jul-Aug 87.
Journal Announcement: GRAI8810; STAR2607

Shaskan, Edward G.; Dolinsky, Zelig S.
Elevated Endogenous Breath Acetaldehyde Levels among Abusers of Alcohol and Cigarets.
Location: Health Cent., Univ. Connecticut, Farmington, CT, 06032, USA
Journal: Prog. Neuro-Psychopharmacol. Biol. Psychiatry Date: 1985
Volume: 9 Number: 3 Pages: 267-72

Shepson, Paul B.; Kleindienst, Tadeusz E.; Edney, Edward O.;
Cupitt, Larry T.; Claxton, Larry D.
Acetaldehyde: The Mutagenic Activity of its Photooxidation Products.
Northrop Services Inc, Research Triangle Park, NC, USA
Source: National Meeting - American Chemical Society, Division of Environmental Chemistry 191st v 26 n 1. Publ by ACS, Washington, DC, USA
p27-29
Publication Year: 1986

Stowell, Allan; Johnsen, Jon; Aune, Haakon; Vatne, Karleif;
Ripel, Aase; Moerland, Joerg
A Reinvestigation of the Usefulness of Breath Analysis in the
Determination of Blood Acetaldehyde Concentrations.
Location: Natl. Inst. Forensic Toxicol., Oslo, Norway
Journal: Alcohol.: Clin. Exp. Res. Date: 1984 Volume: 8 Number: 5
Pages: 442-7

Tornqvist, M.; Ostermangolkar, S.; Kautiainen, A.; Naslund, M.;
Calleman, C. J.; Ehrenberg, L.
Methylations in Human-Hemoglobin.
Univ Stockholm, Dept Radiobiol/S-10691 Stockholm//Sweden/
Mutation Research, 1988, v204, n3, p521-529

Torres, J.
Genetic Toxicity Studies of Organic Chemicals Found as Contaminants in
Spacecraft Cabin Atmospheres.
Southeastern Louisiana Univ., Hammond. Dept. of Biological Sciences.
Corp. Source Codes: 056586002; U5944451
Sponsor: National Aeronautics and Space Administration, Washington, DC.
Jun 87 14p
In NASA. Lyndon B. Johnson Space Center, National Aeronautics and Space
Administration (NASA)/American Society for Engineering Education
(ASEE)
Summer Faculty Fellowship Program, v2 14p 1986.
Journal Announcement: GRAI8721; STAR2519

United States Environmental Protection Agency
Health Assessment Document for Acetaldehyde. External Review Draft.
Location: Environ. Criter. Assess. Off., Research Triangle Park, NC, USA
Journal: Report Date: 1987 Number: EPA/600/8-86/015A; Order No.
PB87-202446 Pages: 216 pp.

Yan, R. T.; Zhu, C. X.; Chen, J. S.
Oxidation Product(s) in Acetaldehyde Reacts with NAD(P)H and Interferes
with Assay of Alcohol Dehydrogenase.
Department of Anaerobic Microbiology, Virginia Polytechnic Institute,
Blacksburg 24061.
Anal Biochem (United States) Aug 1 1987, 164 (2) p362-6,
ISSN 0003-2697 Journal Code: 4NK

Zaleska, M. M.; Gessner, P. K.
Metabolism of [14C]Paraldehyde in Mice in Vivo, Generation and Trapping
of Acetaldehyde.
J Pharmacol Exp Ther Mar 1983, 224 (3) p614-9, ISSN 0022-3565
Journal Code: JP3

Attachment 3
February 1989

State of California
AIR RESOURCES BOARD

Procedure for Handling Trade Secret Information

The information that you provide to the Air Resources Board may be released (1) to the public upon request, except trade secrets which are not emission data or other information which is exempt from disclosure or the disclosure of which is prohibited by law, and (2) to the federal Environmental Protection Agency, which protects trade secrets as provided in Section 114(c) of the Clean Air Act and amendments thereto (42 USC 7401 et seq.) and in federal regulations and (3) to other public agencies provided that those agencies preserve the protections afforded information which is identified as a trade secret, or which is otherwise exempt from disclosure by law (Section 91000 Title 17, California Administrative Code and California Health and Safety Code Section 39660).

Trade secrets, which are defined in California Government Code Section 6254.7, are not considered public records and therefore will not be released to the general public. However, the Public Records Act (Government Code Section 6250 et seq.) provides that air pollution emission data are always public records, even if the data come within the definition of trade secrets. On the other hand, the information used to calculate air pollution emissions data may be withheld from the public if the information is a trade secret.

If you believe that any of the information you are providing is a trade secret or otherwise exempt from disclosure under any other provision of the law, you should identify it as such at the time of submission (Health and Safety Code Section 39660 (e)). The Air Resources Board may ask you to provide documentation of your claim of trade secret of exemption at a later date.

Chevron

Chevron Environmental Health Center, Inc.
A Subsidiary of Chevron Corporation
15299 San Pablo Avenue, Richmond, California
Mail Address: P.O. Box 4054, Richmond, CA 94804-0054

April 11, 1989

Robert Barham, Chief
Toxic Air Contaminant Identification Branch
Re: Acetaldehyde
California Air Resources Board
P.O. Box 2815
Sacramento, California 95812

Dear Mr. Barham:

Thank you for the opportunity to provide input into your review of acetaldehyde. We do not have any in-house data on the possible health effects of acetaldehyde, but we suggest that the following useful citation be included in your list of references.

Egle, John L., Jr.
Retention of Inhaled Acetaldehyde in Man
Journal of Pharmacol. Expt'l. Therapeutics 174 (1):14-19, 1970.

If you have any questions, please don't hesitate to contact me at (415) 231-6088.

Sincerely,

Thomas F. Booze, Ph.D.
Toxicologist

TFB:temp2-c/0489-025

Pacific Gas and Electric Company

P.O. Box 7640
San Francisco, CA 94120
415/972-6901
Telecopy 415/972-9201

John F. McKenzie
Director
Environmental Planning

April 28, 1989

Robert Barham, Chief
Toxic Air Contaminant Identification Branch
Re: Acetaldehyde
California Air Resources Board
P.O. Box 2815
Sacramento, CA 95812

Dear Mr. Barham:

Request for Information on
Public Health Effects of Acetaldehyde

Pacific Gas and Electric Company (PG&E) has reviewed your March 24, 1989 bibliography on the health effects of acetaldehyde. We have no additional information relevant to human exposure at this time.

PG&E appreciates this opportunity to be invited to provide information regarding acetaldehyde. To assist ARB in implementing the toxic air contaminant program, PG&E would appreciate being requested to provide information for other candidate substances in the future. If you have any other questions, please call me or Kee Tsang at (415) 972-6911.

Sincerely,

nassco

NATIONAL STEEL AND SHIPBUILDING COMPANY

A MORRISON-KNUDSEN COMPANY

March 30, 1989

Robert Barham, Chief
Toxic Air Contaminent Identification Branch
Re: Acetaldehyde
California Air Resources Board
P. O. Box 2815
Sacramento, Ca. 95812

Dear Robert,

I am responding to your letter dated March 24, 1989, requesting information regarding acetaldehyde. I have reviewed our chemical inventory lists and discussed the request with our Industrial Hygienist and at this time we do not use acetaldehyde in any of our operations within the shipyard.

Sincerely,

Michael Chee
Sr. Facilities Engineer
NASSCO STEEL & SHIPBUILDING
P. O. Box 85278
San Diego, Ca. 92138

HARBOR DR. AND 28TH STREET. P O. BOX 85278. SAN DIEGO, CA 92138

UNITED NATIONS ENVIRONMENT PROGRAMME
PROGRAMME DES NATIONS UNIES POUR L'ENVIRONNEMENT

International Register
of Potentially Toxic
Chemicals

IRPTC/RISCPT

Le Registre international
des substances chimiques
potentiellement toxiques

QR 2124

18 April 1989

Avec les compliments
du

With the compliments
of

Meera Laurijssen (Mrs)
Asst. Scientific Affairs Officer
IRPTC/UNEP

Further to your letter of 24 March 1989 please find enclosed the IRPTC data profile on Acetaldehyde.

Mr Peter D. Venturini
Chief, Stationary Source Div.
Air Resources Board
1102 Q Street
Sacramento, CA. 95812
USA

Bureaux/Office location: 16, avenue Jean-Trembley, Petit-Saconnex, 1209 Genève
Telephone: 94 84 00 - 98 58 50 - Telex 28 877 UNEP-CH - Cables: UNITERRA, GENEVA

UNITED NATIONS ENVIRONMENT PROGRAMME

IRPTC

International Register of Potentially Toxic Chemicals
United Nations Environment Programme
IRPTC/UNEP, Palais des Nations, 1211 Genève 10, Suisse

QR 2124

APRIL 1989

IRPTC DATA PROFILE ON : ACETALDEHYDE

```
AB1925000  ACETALDEHYDE
75-07-0

EVALUATION-APPRAISAL:
THERE IS SUFICIENT EVIDENCE FOR THE CARCINOGENICITY OF ACETALDEHYDE TO EXPERIMENTAL ANIMALS.
THERE IS INADEQUATE EVIDENCE FOR THE CARCINOGENICITY OF ACETALDEHYDE TO HUMANS.

IARMB8 MONOGRAPHS ON THE EVALUATION OF THE CARCINOGENIC RISK OF CHEMICAL 36,120,1985
       $ TO MAN                                                       IARC
----                                                                                          REC:  912

-----------------------------------------------------------------------

EVALUATION-APPRAISAL:
CLASSIFIED AS "POSSIBLY CARCINOGENIC TO HUMANS".

IARMB8 MONOGRAPHS ON THE EVALUATION OF THE CARCINOGENIC RISK OF CHEMICAL SUPPL.7,32,1987
       $ TO MAN                                                       IARC
----                                                                                          REC:  1396

-----------------------------------------------------------------------

- - - - T E S T   D E S C R I P T I O N - - - -        - - - - T E S T   C O N D I T I O N S - - - -

ORGANISM:            HAMST                             STUDY TYPE:         LAB
LIFESTAGE,AGE:       6WK
SEX:                 F             M
NUMBER EXPOSED:      36/DOSE       36/DOSE
NUMBER CONTROL:      36            36
EXPOSURE
 ROUTE:      IHL
 DOSE:       2970 - 4500 MG/M3 AIR
 TYPE:       LONG
 PERIOD:     52WK
 FREQUENCY:  7H/D            5D/WK

SPECIES-STRAIN-SYSTEM:
SYRIAN GOLDEN
EXPOSURE:
CONCENTRATION DECREASED FORM 4500 TO 2970 MG/M3
TEST METHODS:
6 ANIMALS WERE KILLED AND EXAMINED FROM EACH GROUP AT WEEK 52.  THE REMAINING WERE OBSERVED TILL 81 WEEK AND KILLED
- - - - E F F E C T - - - -

ORGAN EFFECT REV      ONSET     SEX    AFFECTED  IN    COMMENTS ON RESULT
                                     EXPOSED - CONTROLS

      NEF                                              6 ANIMALS FROM EACH GROUP HAD NO TUMOUR
PULMO NEO                        M  8/29      0/30     RESPIRATORY TRACT TUMOUR (P<0.05)
PULMO NEO                        F  5/29      0/28     RESPIRATORY TRACT TUMOUR (P<0.05)
PULMO CAR                                              COMPOUND-INDUCED TUMORS WERE PREDOMINANTLY LARYNGEAL CARCINO
      NEO                                              MAS WITH A FEW LARYNGEAL POLYPS, & NASAL POLYPS & CARCINOMAS
```

CHEMICAL NAME: ACETALDEHYDE
COMMON NAME: ACETALDEHYDE

RTECS RN: AB1925000
CAS RN: 75-07-0

REC: 333

```
AB1925000   ACETALDEHYDE
75-07-0

EVALUATION-APPRAISAL:
NO DATA WERE AVAILABLE ON THE GENETIC AND RELATED EFFECTS OF ACETALDEHYDE IN HUMANS.

IIARM88 MONOGRAPHS ON THE EVALUATION OF THE CARCINOGENIC RISK OF CHEMICAL SUPPL.7,77,1987
        S TO MAN                                                          IARC
                                                                                                    REC:   965
----

-----------------------------------------------------------------------------

EVALUATION-APPRAISAL:
NO DATA WERE AVAILABLE ON THE GENETIC AND RELATED EFFECTS OF ACETALDEHYDE IN HUMANS.
ACETALDEHYDE INCREASED THE INCIDENCE OF SISTER CHROMATID EXCHANGES IN BONEMARROW CELLS OF MICE AND HAMSTER TREATED IN VI
VO AND INDUCED CHROMOSOMAL ABERRATIONS IN RAT EMBRYOS EXPOSED IN VIVO. IT INDUCED DNA CROSS-LINKS, CHROMOSOMAL ABERRATIO
NS AND SISTER CHROMATID EXCHANGES IN HUMAN CELLS IN VITRO AND CHROMOSOMAL ABERRATIONS, MICRONUCLEI AND SISTER CHROMATID
EXCHANGES IN CULTURED RODENT CELLS. IT INDUCED CHROMOSOMAL ABERRATIONS, MICRONUCLEI AND SISTER CHROMATID EXCHANGES IN PL
ANTS AND DNA DAMAGE AND MUTATION IN BACTERIA. ACETALDEHYDE INDUCED CROSSLINKS IN ISOLATED DNA.
EVALUATION BASED ON IARC MONOGRAPH 36, 1985.

IIARM88 MONOGRAPHS ON THE EVALUATION OF THE CARCINOGENIC RISK OF CHEMICAL SUPPL. 6,21,1987
        S TO MAN
                                                                                                    REC:  1340
----

-----------------------------------------------------------------------------
```

A

```
IIARMB8 MONOGRAPHS ON THE EVALUATION OF THE CARCINOGENIC RISK OF CHEMICAL 36,101,1985
        S TO MAN
EJCODS  EUROPEAN JOURNAL OF CANCER & CLINICAL ONCOLOGY                   18,13,1982
                                                                         FERON, V.J. ET AL                         REC:   669

--------------------------------------------------------------------------
```

AB1925000 ACETALDEHYDE
75-07-0

TREATMENT & DISPOSAL METHODS:	FROM THIS REFERENCE	RECOMMENDED	RECOMMENDABLE	NOT RECOMMENDABLE
			INCINERATION	

DISSOLVE IN A COMBUSTIBLE SOLVENT SUCH AS ALCOHOL OR BENZENE, THEN SPRAY THE SOLUTION INTO THE FURNACE WITH AFTERBURNER.

ITIIT* TOXIC AND HAZARDOUS INDUSTRIAL CHEMICALS SAFETY MANUAL FOR HANDLI -,1,1979
NG AND DISPOSAL WITH TOXICITY AND HAZARD DATA.

REC: 346

```
AB1925000   ACETALDEHYDE
75-07-0

MEDIUM       ANALYTICAL METHOD    DETECTION LIMIT       SAMPLE SIZE      COMMENTS

AIR          HPLC-UV (245NM)      LESS THAN 0.325UG     60L

COMMENTS:
AIR IS DRAWN THROUGH A MIDGET BUBBLER CONTAINING A BUFFERED SOLUTION OF GIRARD T REAGENT. THE ACETALDEHYDE-GIRARD T DERI
VATIVE IS THEN ANALYSED. THE METHOD WAS VALIDATED OVER THE RANGE OF 170-670MG/M3 AT T.= 21C AND P.= 756MM HG.
HPLC EQUIPPED WITH A VARIABLE WAVELENGTH UV DETECTOR SET AT 245NM AND A SAMPLE INJECTION VALUE WITH A 50UL EXTERNAL SAMP
LE LOOP.
COLUMN PACKED WITH ZIPAX SCX- STAINLESS STEEL CONDITIONS: COLUMN T.= AMBIENT. FLOW RATE = 0.75ML/MI, MOBILE PHASE =
0.022M NA2HPO4 AND 0.019M NAH2PO4 IN 20% ETHANOL.

----
NSHAM*  NIOSH MANUAL OF ANALYTICAL METHODS                    5,S-345,1979                    REC:   402
```

AB1925000 ACETALDEHYDE
75-07-0

AREA	TYPE	SUBJECT		DESCRIPTOR	LEVELS,REMARKS AND REFERENCE	
ARG	REG	AIR	OCC	MPC	TWA : 180 MG/M3 (100 PPM), STEL : 270 MG/M3 (150 PPM) ARGENTINIAN LEGISLATION, DECRETO. , 351,-,1979	EFFECTIVE DATE: 22MAY1979 ENTRY DATE IN IRPTC: MAY1985
AUS	REC	AIR	OCC	TLV	TWA: 180MG/M3 (100PPM) STEL: 270MG/M3 (150PPM) APPROVED OCCUPATIONAL HEALTH GUIDE THRESHOLD LIMIT VALUES , -,-,1983	EFFECTIVE DATE: ENTRY DATE IN IRPTC: MCH1985
BEL	REC	AIR	OCC	TLV	TWA: 180MG/M3 (100PPM); STEL: 270MG/M3 (150 PPM). THRESHOLD LIMIT VALUES (VALEURS LIMITES TOLERABLES) , -,-,1984	EFFECTIVE DATE: ENTRY DATE IN IRPTC: JUL1987
CAN	REG	AIR	OCC	TLV	TWA: 100PPM (180MG/M3). STEL: 150PPM (270MG/M3) CONSOLIDATED REGULATIONS OF CANADA , 1978*10(997),7749,1978	EFFECTIVE DATE: 24APR1980 ENTRY DATE IN IRPTC: SEP1986
CAN	REG	TRNSP LABEL PACK		CLASS RQR ,	PIN (PRODUCT IDENTIFICATION NO.): 1089. CLASS (3.1): FLAMMABLE LIQUID WITH FLASHPOINT LESS THAN -18C. CLASS (9.2): HAZARDOUS TO THE ENVIRONMENT. SPECIAL PROVISIONS: (46),(56),(90),(99). PACK GROUP I (I=GREAT DANGER, III=MINOR DANGER). MAXIMUM AMOUNT PER PACKAGE THAT MAY BE TRANSPORTED ON A PASSENGER AIRCRAFT, TRAIN OR ROAD VEHICLE :PROHIBITED. MAXIMUM AMOUNT PER PACKAGE THAT MAY BE TRANSPORTED ON A CARGO AIRCRAFT: 30L (CANADA GAZETTE PART II 1985*119(77),393,1985 AS LAST AMENDED BY THE REFERENCE GIVEN. THE REGULATIONS FALL UNDER THE TRANSPORTATION OF DANGEROUS GOODS ACT). CANADA GAZETTE PART II , 1987*121(13),2304,1987	EFFECTIVE DATE: 11JUN1987 ENTRY DATE IN IRPTC: MAY1988
CHE	REG	AIR	OCC	MAK	TWA: 90MG/M3 (50PPM) ZULAESSIGE WERTE AM ARBEITSPLATZ (PERMITTED VALUES IN THE WORKPLACE) , -,-,1987	EFFECTIVE DATE: ENTRY DATE IN IRPTC: DEC1987
CSK	REG	AIR	OCC	MAC	TWA = 200.0 MG/M3 ; CLV = 400.0 MG/M3 HYGIENICKE PREDPISY MINISTERSTVA ZDRAVOTNICTVI CSR (HYGIENIC REGULATIONS OF MINISTRY OF HEALTH OF CSR) , 58,-,1985	EFFECTIVE DATE: MCH1985 ENTRY DATE IN IRPTC: AUG1985
DDR	REG	AIR	OCC	MAC	STEL: 100MG/M3 (IRRITANT TO THE RESPIRATORY SYSTEM). MAXIMALE ZULAESSIGE KONZENTRATIONEN GESUNDHEITSGEFAERDENDER STOFFE IN DER LUFT AM ARBEITSPLATZ (MAXIMUM ALLOWABLE CONCENTRATIONS OF NOXIOUS SUBSTANCES IN THE AT , -,1,1987	EFFECTIVE DATE: ENTRY DATE IN IRPTC: APR1988

```
DEU      REC  AIR    OCC    MAK     8H-TWA: 90MG/M3 (50 PPM); 5MI-STEL: 180MG/M3     EFFECTIVE DATE:
                                    (100PPM) CEILING VALUE, 8X/SHIFT. LOCAL IRRIT    ENTRY DATE IN IRPTC: SEP1987
                                    ANT. (VP=100KPA AT 20C).
                                    SUBSTANCE SUSPECT OF CARCINOGENIC POTENTIAL.
                                    THE MAK VALUE IS THEREFORE TENTATIVELY RETAIN
                                    ED.
                                    DEUTSCHE FORSCHUNGSGEMEINSCHAFT, "MAXIMALE ARBEITSPLATZKONZENTRAT
                                    IONEN UND BIOLOGISCHE ARBEITSSTOFFTOLERANZWERTE"
                                    (MAXIMUM WORKSITE CONCENTRATION AND BIOLOGICAL TOLERANCE VALUES , XXIII,15,1987

DEU      REG  AIR    EMI    MXL     THE SUBSTANCE BELONGS TO CLASS I                 EFFECTIVE DATE: 28FEB1986
                                    THE AIR EMISSIONS OF ORGANIC COMPOUNDS MUST N    ENTRY DATE IN IRPTC: JUN1986
                                    OT EXCEED (AS THE SUM OF ALL COMPOUNDS IN ONE
                                     CLASS) THE FOLLOWING MASS CONCENTRATIONS:
                                    CLASS I: 20MG/M3 AT A MASS FLOW OF >=0.1KG/H
                                    CLASS II: 100MG/M3 AT A MASS FLOW OF >=2KG/H
                                    CLASS III: 150MG/M3 AT A MASS FLOW OF >=3KG/H
                                    IF COMPOUNDS FROM DIFFERENT CLASSES ARE PRESE
                                    NT THE MASS CONCENTRATION MUST NOT EXCEED
                                    0.15G/M3.
                                    GEMEINSAMES MINISTERIALBLATT
                                    (JOINT MINISTERIAL PAPERS) , 7,93-143,1986

DEU      REC  AQ            CLASS   THIS SUBSTANCE IS CLASSIFIED AS HARMFUL TO       EFFECTIVE DATE:
              STORE                 WATER (CLASS 2). (THE DIFFERENT CLASSES ARE:     ENTRY DATE IN IRPTC: APR1986
              TRNSP                 CLASS 0 = IN GENERAL NOT HARMFUL; CLASS 1 =
                                    SLIGHTLY HARMFUL; CLASS 2 = HARMFUL; CLASS 3
                                    = VERY HARMFUL).
                                    IN ORDER TO PROTECT WATER, APPROPRIATE SECURI
                                    TY MEASURES SHOULD BE TAKEN ACCORDINGLY DURIN
                                    G STORAGE, LOADING AND TRANSPORT.
                                    GEMEINSAMES MINISTERIALBLATT
                                    (JOINT MINISTERIAL PAPERS) , 11(G3191A),173,1985

DEU      REG  LABEL -       RQR     FOR LABEL AND PACKAGING SEE EEC (OJEC L106,      EFFECTIVE DATE: 1OCT1980
              PACK  -       RQR     18,1982).                                        ENTRY DATE IN IRPTC: JUN1986
              USE    OCC    RSTR    HANDLING OF SOME GROUPS OF CHEMICALS (INCL.
                                    CARCINOGENIC, MUTAGENIC, POISONOUS, TERATOGEN
                                    IC, EXPLOSIVE AND EASILY INFLAMMABLE COMPOUND
                                    S) IS PROHIBITED OR RESTRICTED FOR PREADULTS
                                    AND PREGNANT OR NURSING WOMEN.
                                    (BGBL I.S.1071,1536/2159,1980, AND BGBL I.S.
                                    2069,1980 AS LAST AMENDED BY THE REFERENCE
                                    GIVEN).
                                    BUNDESGESETZBLATT , I.S.140,140,1982

EEC      REG  LABEL         RQR     F - FLAMMABLE. XI - IRRITANT. EXTREMELY FLAMM    EFFECTIVE DATE:
              PACK          RQR     ABLE (R 12). IRRITATING TO EYES AND RESPIRATO    ENTRY DATE IN IRPTC: APR1986
                                    RY SYSTEM (R 36/37). KEEP CONTAINER IN A WELL
                                    -VENTILATED PLACE (S 9). KEEP AWAY FROM SOURC
                                    ES OF IGNITION - NO SMOKING (S 16). DO NOT EM
                                    PTY INTO DRAINS (S 29). TAKE PRECAUTIONARY ME
                                    ASURES AGAINST STATIC DISCHARGES (S 33).
                                    OFFICIAL JOURNAL OF THE EUROPEAN COMMUNITIES , L.106,18,1982
```

FIN	REC	AIR	OCC	MPC	TWA: 90MG/M3 (50PPM) STEL: 135MG/M3 (75PPM) LUFTFOEORENINGAR PAA ARBETSPLATSEN (AIR POLLUTANTS AT THE WORKPLACE) , -,-,1982	EFFECTIVE DATE: ENTRY DATE IN IRPTC: MCH1985
GBR	REC	AIR	OCC	RECL	8H-TWA: 180MG/M3 (100PPM). STEL(10MI-TWA): 270MG/M3 (150PPM) GUIDANCE NOTE FROM THE HEALTH AND SAFETY EXECUTIVE , EH40/87,8,1987	EFFECTIVE DATE: ENTRY DATE IN IRPTC: JUN1987
GBR	REG	AIR	EMI	RQR	PROVISIONS HAVE BEEN MADE WITH A VIEW TO CONTROLLING THE EMISSION INTO THE ATMOSPHERE OF NOXIOUS OR OFFENSIVE SUBSTANCES FROM ALDEHYDE WORKS IN WHICH FORMALDEHYDE, ACETALDEHYDE OR ACROLEIN OR THE METHYL, ETHYL OR PROPYL DERIVATIVES OF ACROLEIN ARE MADE. STATUTORY INSTRUMENTS , 943,-,1983	EFFECTIVE DATE: 5AUG1983 ENTRY DATE IN IRPTC: APR1986
GBR	REG	AIR	EMI	RQR	IT SHALL BE THE DUTY OF THE PERSON HAVING CONTROL OF ANY PREMISES OF A CLASS PRESCRIBED ELSEWHERE UNDER THIS REFERENCE TO USE THE BEST PRACTICABLE MEANS FOR PREVENTING THE EMISSION INTO THE ATMOSPHERE FROM THE PREMISES OF NOXIOUS OR OFFENSIVE SUBSTANCES SUCH AS ACETALDEHYDE AND FOR RENDERING HARMLESS AND INOFFENSIVE SUCH SUBSTANCES AS MAY BE SO EMITTED. (APPLIES TO ACETALDEHYDES). STATUTORY INSTRUMENTS , 943,1,1983	EFFECTIVE DATE: 5AUG1983 ENTRY DATE IN IRPTC: APR1986
GBR	REG	TRNSP LABEL		CLASS RQR	LABELLING OF ROAD TANKERS : FLAMMABLE LIQUID. , EMERGENCY ACTION CODE : 2YE STATUTORY INSTRUMENTS , 1702,1,1978	EFFECTIVE DATE: 28MCH1979 ENTRY DATE IN IRPTC: JAN1983
GBR	REG	LABEL PACK USE	- - OCC	RQR RQR RSTR	FOR LABEL AND PACKAGING, SEE EEC (OJEC** L.106,18,1982) STATUTORY INSTRUMENTS , 17,1,1983	EFFECTIVE DATE: 1SEP1978 ENTRY DATE IN IRPTC: APR1986
HUN	REG	AIR	OCC	MAC	TWA: 50MG/M3; STEL(30 MIN): 250MG/M3 HUNGARIAN STANDARD MSZ NO. , 21461-78,-,1978	EFFECTIVE DATE: ENTRY DATE IN IRPTC: MCH1985
ITA	REC	AIR	OCC	TLV	100MG/M3 (55PPM) VALORI LIMITE PONDERATI (THRESHOLD LIMIT VALUES) , -,-,	EFFECTIVE DATE: ENTRY DATE IN IRPTC: MCH1985
JPN	REC	AIR	EMI	MXL	EMISSION STANDARD BECAUSE OF OFFENSIVE ODOR : 0.05-0.5PPM (0.09-0.9 MG/M3) QUALITY OF THE ENVIRONMENT IN JAPAN , -,396,1981	EFFECTIVE DATE: ENTRY DATE IN IRPTC: DEC1982
NLD	REC	AIR	OCC	MXL	TWA: 180MG/M3 (100PPM) NATIONALE MAC-LIJST (NATIONAL MAC-LIST) , -,-,1986	EFFECTIVE DATE: ENTRY DATE IN IRPTC: JUN1987
POL	REG	AIR	OCC	MPC	TWA: 5MG/M3	EFFECTIVE DATE: ENTRY DATE IN IRPTC: MCH1985

A

```
                                  ORDINANCE OF THE MINISTER OF LABOUR, WAGES AND SOCIAL AFFAIRS , 22DEC,-,1982
                                                                                     EFFECTIVE DATE:
ROM   REG  AIR   OCC   MPC        TWA: 100MG/M3; CLV: 200MG/M3                       ENTRY DATE IN IRPTC: MCH1985
                                  ORDINANCE OF THE MINISTRY OF HEALTH , 60,-,1975
                                                                                     EFFECTIVE DATE: JAN1977
SUN   REG  AIR   OCC   MAC        CLV : 5.0 MG/M3 (VAPOUR)                           ENTRY DATE IN IRPTC: JUN1982
                                  GOSUDARSTVENNYI STANDART SSSR
                                  (STATE STANDARD OF USSR) , 12.1.005,-,1976
                                                                                     EFFECTIVE DATE: AUG1984
SUN   REG  AIR   AMBI  MAC        0.01 MG/M3 1X/D, 0.01 MG/M3 AV/D                   ENTRY DATE IN IRPTC: SEP1985
                                  PREDELNO DOPUSTIMYE KONTSENTRATSII (PDK) ZAGRYAZNYAYUSHCHIKH VESH
                                  CHESTV V ATMOSFERNOM VOZDUKHE NASELENNYKH MEST
                                  (MAXIMUM ALLOWABLE CONCENTRATIONS (MAC) OF CONTAMINANTS IN THE , 3086-84,-,1984
                                                                                     EFFECTIVE DATE: OCT1983
SUN   REG  AQ    SURF  MAC        0.2 MG/L                                           ENTRY DATE IN IRPTC: SEP1985
                                  PREDELNO DOPUSTIMYE KONSENTRATSII (PDK) I ORIENTIROVOCHMYE BEZOPA
                                  SNYE UROVNI VOZDEISTVIA (OBUV) UREDMYKH VESHCHESTV V VODE VODNYKH
                                  OBIEKTOV KHOZIAISTVENNO-PITIEVOGO I KULTURNO-BYTOVOGO VODOPOLZOV , 2932-83,-,1983
                                                                                     EFFECTIVE DATE: FEB1985
SUN   REG  SOIL        MAC        10.0 MG/KG                                         ENTRY DATE IN IRPTC: OCT1985
                                  PREDELNO DOPUSTIMYE KONTSENTRATSII (PDK) KHIMICHESKIKH VESHCHESTV
                                  V POCHVE
                                  (MAXIMUM ALLOWABLE CONCENTRATIONS (MAC) OF CHEMICAL SUBSTANCES IN , 3210-85,-,1985
                                                                                     EFFECTIVE DATE: 01JUL1988
SWE   REG  AIR   OCC   HLV        TWA: 45MG/M3 (25PPM), STEL: 90MG/M3 (50PPM).       ENTRY DATE IN IRPTC: FEB1988
                                  ARBETARSKYDDSSTYRELSENS FOERFATTNINGSSAMLING , 1987:12,13,1987
                                                                                     EFFECTIVE DATE: 23NOV1983
SWE   REG  CLASS       CLASS      GIVEN AS AN EXAMPLE OF A HAZARDOUS SUBSTANCE       ENTRY DATE IN IRPTC: OCT1984
                                  STATENS NATURVARDSVERKS FORFATTNINGSSAMLING , 1983:6,1,1983
                                                                                     EFFECTIVE DATE:
USA   REC  AIR   OCC   TLV        TWA: 180MG/M3(100PPM); STEL: 270MG/M3(150PPM)      ENTRY DATE IN IRPTC: OCT1987
                                  AMERICAN CONFERENCE OF GOVERNMENT INDUSTRIAL HYGIENISTS , 1987-88,11,1987
                                                                                     EFFECTIVE DATE:
USA   REG  AIR   OCC   PEL        TWA : 360 MG/M3 (200 PPM)                          ENTRY DATE IN IRPTC: DEC1987
                                  CODE OF FEDERAL REGULATIONS , 29(1910),653,1986
                                                                                     EFFECTIVE DATE:
USA   REG  AQ    EMI   CLASS      THE SUBSTANCE, INCLUDING ANY ISOMERS, HYDRATE      ENTRY DATE IN IRPTC: MAY1986
                                  S, AND SOLUTIONS AND MIXTURES CONTAINING THE
                                  SUBSTANCE, IS DESIGNATED AS A HAZARDOUS SUBST
                                  ANCE FOR PURPOSES OF DISCHARGE (INCLUDING SPI
                                  LLING, LEAKING, ETC.) UNDER THE WATER POLLUTI
                                  ON CONTROL ACT.
                                  CODE OF FEDERAL REGULATIONS , 40(116),33,1981
                                                                                     EFFECTIVE DATE:
USA   REG  AQ    EMI   RQR        UNLESS IN COMPLIANCE WITH A SPECIFIED PERMIT       ENTRY DATE IN IRPTC: MAY1986
                       RSTR       OR PROCEDURE, OWNERS/OPERATORS OF VESSELS OR
                                  ON- OR OFFSHORE FACILITIES MUST NOTIFY THE US
                                  GOVERNMENT OF ANY DISCHARGE OF THIS SUBSTANCE
                                  IN OR ON NAVIGABLE WATERS, ADJOINING SHORELIN
                                  ES OR THE CONTIGUOUS ZONE IN AN AMOUNT EQUAL
                                  TO OR GREATER THAN 454KG IN ANY 24-HOUR
                                  PERIOD. SUCH A DISCHARGE IS A VIOLATION OF
                                  THE FEDERAL WATER POLLUTION CONTROL ACT.
```

A

CODE OF FEDERAL REGULATIONS , 40(117),44,1981

USA	REG	AQ	EMI WASTE	RQR	PERMITS ARE REQUIRED FOR DISCHARGE OF ANY POLLUTANT FROM ANY POINT SOURCE INTO U.S.WATERS. THE FOLLOWING APPLICANTS MUST REPORT QUANTITATIVE DATA FOR THE SUBSTANCE IN THEIR APPLICATIONS: 1) EVERY APPLICANT WHO HAS REASON TO BELIEVE THAT THE SUBSTANCE IS CONTAINED IN ANY OF HIS OUTFALLS INSPECTION, MONITORING AND REPORTING REQUIREMENTS AFTER THE ISSUE OF THE PERMIT ARE SPECIFIED. EVEN IF NOT REQUIRED IN THE PERMIT, DISCHARGE OF THE SUBSTANCE MUST BE REPORTED IF IT EXCEEDS THE HIGHEST OF THE FOLLOWING LEVELS: A) 100UG/L B) 5 TIMES THE MAXIMUM CONC. REPORTED IN THE APPLICATION C) THE LEVEL ESTABLISHED BY THE EPA DIRECTOR. CODE OF FEDERAL REGULATIONS , 40(122.53),121,1981	EFFECTIVE DATE: ENTRY DATE IN IRPTC: MAY1986
USA	REG	AQ TRNSP	EMI	RQR	WHEN CERTAIN SPECIFIED VESSELS CARRYING ACETALDEHYDE IN BULK ARE BOUND FOR OR DEPARTING FROM U.S. PORTS, THE CAPTAIN OF THE PORT MUST BE NOTIFIED AT LEAST 24 HOURS IN ADVANCE AND BE GIVEN CERTAIN INFORMATION. CODE OF FEDERAL REGULATIONS , 33(161),813,1981	EFFECTIVE DATE: ENTRY DATE IN IRPTC: MAY1986
USA	REG	FOOD FEED		ARL	EXEMPTIONS FROM RESIDUE TOLERANCE REQUIREMENTS IS MADE FOR THIS FUNGICIDE, WHEN USED POSTHARVEST AS A STORAGE FUMIGANT, IN OR ON CERTAIN SPECIFIED PLANT PRODUCTS. CODE OF FEDERAL REGULATIONS , 40(180),269,1981	EFFECTIVE DATE: ENTRY DATE IN IRPTC: APR1986
USA	REG	TRNSP		RSTR RQR CLASS	ACETALDEHYDE, WHEN CARRIED IN BULK, IS CLASSIFIED AS A 'CARGO OF PARTICULAR HAZARD' FOR PURPOSES OF REGULATIONS GOVERNING HANDLING OF DANGEROUS CARGOES IN OR CONTIGUOUS TO WATER-FRONT FACILITIES. A PERMIT IS REQUIRED FOR HANDLING SUCH CARGO. OFFICIAL JOURNAL OF THE EUROPEAN COMMUNITIES , 33(126),613,1981	EFFECTIVE DATE: ENTRY DATE IN IRPTC: APR1986
USA	REG	USE	OCC	RSTR	NO POWER OPERATED INDUSTRIAL TRUCK MAY BE USED IN ATMOSPHERES WHERE A HAZARDOUS CONCENTRATION OF THE SUBSTANCE IS PRESENT. CODE OF FEDERAL REGULATIONS , 29(1910),419,1981	EFFECTIVE DATE: ENTRY DATE IN IRPTC: APR1986
USA	REG	WASTE		CLASS RSTR RQR	THE SUBSTANCE, IF IT IS A COMMERCIAL CHEMICAL PRODUCT, IS IDENTIFIED AS A 'TOXIC WASTE' SUBJECT TO REGULATION AND NOTIFICATION REQUIREMENTS. CODE OF FEDERAL REGULATIONS , 40(261),368,1981	EFFECTIVE DATE: 19NOV1980 ENTRY DATE IN IRPTC: MAY1986
USA	REG	WASTE		CLASS	IN ACCORDANCE WITH THE LAW REQUIRING THE PRESIDENT TO IDENTIFY THE 400 HAZARDOUS WASTE FACILITIES IN THE USA WARRANTING THE HIGHEST PRIORITY FOR REMEDIAL ACTION, EPA HAS ESTABLISHED A HAZARD RANKING SYSTEM TO SET PRIORITIES. A CRITERION FOR ESTABLISHING THE RELATIVE RISK OR DANGER IS THE CHARACTERISTICS OF WASTES IN A GIVEN FACILITY. FOR EXAMPLE ACETALDEHYDE IS RATED AS A NON-PERSISTENT COMPOUND.	EFFECTIVE DATE: ENTRY DATE IN IRPTC: JUN1983

FEDERAL REGISTER , 47,31219,1982

USA	REG	WASTE		CLASS	IN ACCORDANCE WITH THE LAW REQUIRING THE PRESIDENT TO IDENTIFY THE 400 HAZARDOUS WASTE FACILITIES IN THE USA WARRANTING THE HIGHEST PRIORITY FOR REMEDIAL ACTION, EPA HAS ESTABLISHED A HAZARD RANKING SYSTEM TO SET PRIORITIES. A CRITERION FOR ESTABLISHING THE RELATIVE RISK OR DANGER IS THE CHARACTERISTICS OF WASTES IN A GIVEN FACILITY. FOR EXAMPLE, FOR ACETALDEHYDE THE TOXICITY VALUE IS 3, THE PERSISTENCE VALUE IS 0, THE IGNITION VALUE IS 3 AND THE REACTIVITY VALUE IS 2. (ALL RATINGS RANGING FROM 0 TO 3) FEDERAL REGISTER , 47,31219,1982	EFFECTIVE DATE: ENTRY DATE IN IRPTC: JUN1983
YUG	REG	AIR	OCC	MAC	TWA: 360MG/M3 (200PPM) ORDINANCE , 24-3698/1,-,1971	EFFECTIVE DATE: ENTRY DATE IN IRPTC: MCH1985

APPENDIX G

HEALTH EFFECTS REQUEST LETTER TO DEPARTMENT OF HEALTH SERVICES AND LETTER OF RESPONSE

State of California

MEMORANDUM

To : Ed Mendoza, Assistant Deputy Director
Public Health
Department of Health Services
714 P Street
Sacramento, CA 95814

Date : June 22, 1989

Subject : Health Effect
Evaluation o
Acetaldehyde

James D. Boyd
Executive Officer
From : Air Resources Board

This memorandum is a formal request that the Department of Health Services evaluate the health effects of acetaldehyde as a candidate toxic air contaminant in accordance with Health and Safety Code Section 39650 *et seq*.

Attached for your staff's consideration in evaluating acetaldehyde are the references on acetaldehyde health effects which were identified in an Air Resources Board (ARB) letter of public inquiry and provided by the public in response to that letter.

Ambient air monitoring of acetaldehyde began earlier this year. Although we do not have specific data at this time, we will provide ambient levels data for your staff later this year.

According to Health and Safety Code Sections 39660-62, the Department of Health Services has 90 days from receipt of this letter to submit a written evaluation to the ARB with recommendations on the health effects of acetaldehyde. If necessary, the Department of Health Services may request a 30 day extension.

If you have questions regarding this request, please contact me at 445-4383 or have your staff contact Robert D. Barham, Chief of the Toxic Air Contaminant Identification Branch at 322-7072.

Attachment

cc: Jananne Sharpless, Chairwoman, ARB
George Alexeeff, DHS, w/Atchs
Richard Jackson, DHS, w/Atchs
Jack Parnell, DFA
Members of the Scientific Review Panel
Assemblywoman Sally Tanner
Senator Ralph Dills
Senator Art Torres

State of California

Office of Environmental Health Hazard Assessment

→ Lon

Memorandum

Date : August 12, 1991

To : Peter Venturini, Chief
Stationary Source Division
Air Resources Board
1102 Q Street
Sacramento, CA 95814

RECEIVED
AUG 16 1991
Stationary Source
Division
Air Resources Board

From : Office of the Director
714 P Street, Room 460
Sacramento, CA 95814
(916) 324-7572

Subject : Health Effects of Acetaldehyde

Attached is the document prepared in response to your request for the assistance of the Office of ~~Environmental Health~~ Hazard Assessment in evaluating the health effects of acetaldehyde as a potential toxic air contaminant.

Steven A. Book, Ph.D.
Acting Director

Attachment

cc: Jim Wells, Acting Director
Department of Pesticide Regulation
1220 N Street
Sacramento, CA 95814

Jananne Sharpless, Chairwoman
Air Resources Board
P. O. Box 2815
Sacramento, CA 95812

The Honorable Sally Tanner
Member of the Assembly
State Capitol, Room 4146
Sacramento, CA 95814

Copies of document can be requested from:

Office of Environmental Health Hazard Assessment
Air Toxicology and Epidemiology Section
2151 Berkeley Way, Annex 11, Room 235
Berkeley, CA 94704

APPENDIX H

GLOSSARY

Glossary

Aldehyde: Any of a class of highly reactive organic chemical compounds obtained by oxidation of primary alcohols, characterized by the common group CHO.

Aliphatic: Organic chemical compounds in which the carbon atoms are linked in open chains rather than rings.

Ambient air: Outdoor air.

Annual averaged concentrations: The yearly mean value of monitored ambient air concentrations from the sum of the monthly means. This is a mechanism to smooth data from seasonal and diurnal variations. Can apply to either one specific site or on a statewide basis.

Annual average population-weighted concentration: The annual average population-weighted exposure divided by the total population exposed.

Annual average population-weighted exposure: The sum of the annual average outdoor concentration (C) estimated for each census tract multiplied by the population (P) exposed in each census tract,

e.g. $E = (C_1 \times P_1) + (C_2 \times P_2)$ ....

Areawide exposure: The average exposure the general population experiences in a large area surrounding sources of TAC's. This is generally based on ambient air monitoring data for populated areas.

Atmospheric Lifetime: The atmospheric lifetime τ of a chemical is defined as the time to decay to a concentration of 1/e (= 0.368) of the initially present concentration of that chemical.

Bioaccumulation: The process by which a chemical becomes more concentrated in an organism than it is in the environment of the organism.

Carcinogenic effects: Effects that cause cancer.

Catalyst: A substance, usually present in small amounts relative to the reactants, that modifies, especially increases, the rate of chemical reaction without being consumed in the process.

Catalytic converter: A reaction chamber typically containing a finely divided platinum-iridium catalyst into which exhaust gases from an automotive engine are passed together with excess air so that carbon monoxide and hydrocarbon pollutants are oxidized to carbon dioxide and water

Catalytic cracking: Using heat, pressure and catalysts to convert heavy oils from distillations, coking and deasphalting processes into lighter products such as gasoline.

Coefficient of variation (CV): The CV is the precision relative to the mean of the population expressed in percentages. The CV is equal to the standard deviation of all observations divided by the mean of all observations multiplied by 100.

Coking operation: Coke is the solid carbon residue obtained from carbon compounds after removal of volatile material by destructive distillation.

Dispersion Modeling: Technique for estimating the annual average concentration of a pollutant resulting from emissions from a specific source. We provide emission rate data to modeling staff which combine these data with data on wind speed and other meteorological conditions impacting the source to estimate the annual average concentration attributable to the source's emissions. The concentration predicted from dispersion modeling is assumed to be in addition to background concentrations measured by ambient air monitoring.

Dose: The concentration of the pollutant and the length of time that the subject is exposed to that pollutant.

Emission factor: An estimate of the rate at which a pollutant is released into the atmosphere as a result of some activity.

Emission rate: The weight of a pollutant emitted per unit of time (e.g. tons/year).

Excess carcinogenic risk: The number of potential excess lifetime cancer cases occurring per million persons continuously exposed for 70 years to a given concentration of a TAC; e.g. the excess carcinogenic risk from acetaldehyde exposure = 7 to 75 potential lifetime cancer cases per million persons continuously exposed to 1 ppb acetaldehyde (7 - 75 potential lifetime cancer cases / 10^6 ppb persons). In this case, 7 would be the risk based on the lower bound potency and 75 would be the risk based on the upper bound potency.

Exposure (E): The concentration (C) of the pollutant in the air multiplied by the population (P) exposed to that concentration over a specified time period, e.g. E = C X P (typical units are millions of ppb-persons averaged over one year).

Humic: Derived from humus. A brown or black organic substance consisting of decayed vegetable matter.

Individual cancer risk: The probability, expressed as chances in a million, that a person experiencing 70 years of continuous areawide outdoor exposure to a TAC will get cancer.

Limit of Detection (LOD): The levels at or above which compound concentrations are not only detectable, but are also quantifiable. LOD = A + 3S : A is the least squares intercept calculated from the multipoint data: S is the standard deviation of replicate determinations of the lowest standard. At least 3 replicates are required. The lowest standard must be run at 1 to 5 times the estimated detection limit. If data is not available in the concentration range near the detection limit, S may be estimated by: S = RSD X A where RSD is the relative standard deviation of the lowest standard analyzed.

Log normal distribution: A random variable's distribution is log normal if the distribution of the log of the variable is a normal distribution.

Mean: A number that represents a set of numbers in any of several ways determined by a rule involving all members of the set; average.

Median: The middle value in a distribution, above and below which lie an equal number of values; midpoint.

Method of Detection: The method used by which the laboratory staff estimates the LOD.

Mutagenic: The ability of a chemical or physical agent to produce inheritable changes in the genetic information stored in the DNA of living cells.

Non-carcinogenic effects: A finite dose, or threshold, below which adverse effects will not occur. Non-cancer effects include birth defects, organ damage, death and many others.

Octanol/water partition coefficient: A laboratory measure of the partitioning of a chemical between n-octanol and water provides Kow, which is related to water solubility

Organic precursor: Carbon compounds that are utilized in the synthesis of further chemicals

Oxidation: The combination of a substance with oxygen. A reaction in which the atoms in an element lose electrons and its valence is correspondingly increased.

Photolysis: Chemical decomposition induced by light or other radiant energy.

Photo-oxidation: Oxidation under the influence of radiant energy (as photochemical oxidation)

Polymer: Any of numerous natural and synthetic compounds of unusually high molecular weight consisting of up to millions of repeated linked units, each a relatively light and simple molecule.

Pyrolysis: Chemical change caused by heat.

Radical: An atom or group of atoms with at least one unpaired electron.

Secondary acetaldehyde: Acetaldehyde formed as a result of photo-oxidation of organic precursors

Sidestream tobacco smoke: Smoke which arises mainly from the passive burning of the cigarette and is released into the environment.

Standard Deviation: A statistic used as a measure of dispersion in a distribution, the square root of the arithmetic average of the squares of the deviations from the mean.

Stationary source: A non mobile source which can be either a point or area source.

Thermal catalytic cracking: Thermal decomposition with catalysis of petroleum to extract low -boiling fractions such as gasoline

Topography: The exact physical configuration of a place or region.

Toxic air contaminant (TAC): An air pollutant which may cause or contribute to an increase in serious illness, or which may pose a present or potential hazard to human health. Substances which have been identified by EPA as hazardous air pollutants (e.g. benzene, asbestos) shall be identified by the Board as toxic air contaminants.

Travel Blank: Used in estimating concentrations of measured substances. A travel blank is a blank system that tells us that our instruments and methods are working properly and that our filled test samples are accurate. It also provides an estimate of the level of contamination due to transport and handling of the cartridges.

Troposphere: The lowest region of the atmosphere between the earths' surface and the tropopause, characterized by decreasing temperature with increasing altitude, about 15 kilometers

Unit Risk: The number of potential excess cancer cases from a lifetime exposure to 1 microgram per cubic meter (ug/m^3) of a given substance; e.g. a unit risk value of 5.5 X 10^{-6} (ug/m^3) would indicate an estimated 5.5 cancer cases per million people exposed to an average concentration of 1 ug/m^3 of a specific carcinogen for 70 years.

Vapor pressure: The pressure exerted by a vapor in equilibrium with its solid or liquid phase.

Volatile: Evaporating readily at normal temperatures and pressures. Capable of being readily vaporized.