He who aims to keep abreast is for ever second best.

Piet Hein, (1905-1996)  
Danish philosopher, mathematician, scientist and author.  
From his book of Grooks.
Introduction

- Longitudinal studies
  - Observational epidemiologic
  - Clinical trials
- Trace developments through my involvement in longitudinal studies over almost four decades
- The conduct of longitudinal studies has been influenced by:
  - Computing
  - Statistics
  - Medicine
The Past

- Seattle Virus Watch – 1965-1971
- The Cauqué Study – 1964-1974
- The Coronary Artery Surgery Study (CASS) – 1973-1983
- Cardiovascular Health Study (CHS) – 1989 - Present
The Seattle Virus Watch Study

- 1965-1969 – PI was John Fox, MD
- Purpose – Understand the natural history of the viruses that infect people and their impact on human health.
- Design
  - Surveillance for two years of families (n=149) with newborn infant for infections with respiratory or enteric viruses
  - Bi-weekly collections of respiratory and fecal specimens from index infant
  - Collection of specimens from family members when illness occurs in family
The Seattle Virus Watch Study

- Study Conduct
  - Data collection
    - 11,500+ person months of observation
    - 3928 illnesses
    - 32,000+ specimens collected
  - Data entry
    - IBM Punch Cards (>300,000 cards)
IBM Punch Card
The Seattle Virus Watch Study

- **Data collection**
  - IBM Punch Cards (>300,000 cards)
  - IBM Port-a-Punch - Data entry in the homes of the participants
    - "do not fold, spindle, or mutilate"
    - “hanging” or “pregnant” chads
  - Machine pre-punched set of cards on enrollment and for each illness

- **Computing – IBM 1130**
  - Early minicomputer
  - First school of public health to have its own computer
The Past –
The Seattle Virus Watch Study

IBM Port-a-Punch
Punch Card Machine
The Seattle Virus Watch Study
IBM 1130 Computer
Results from Virus Watch

SEATTLE VW. 2. DESIGN, OPERATIONS, ILLNESS OBSERVATIONS

**Table 5**

Illness rates and distribution by age and type in continuing observation families

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>All illnesses</th>
<th>% in age group</th>
<th>% of illnesses by type*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total No.</td>
<td>No. per person-year</td>
<td>UR</td>
</tr>
<tr>
<td>&lt;1</td>
<td>653</td>
<td>5.1</td>
<td>16.6</td>
</tr>
<tr>
<td>1</td>
<td>561</td>
<td>5.8</td>
<td>14.3</td>
</tr>
<tr>
<td>2-5</td>
<td>964</td>
<td>5.8</td>
<td>24.6</td>
</tr>
<tr>
<td>6-9</td>
<td>337</td>
<td>3.8</td>
<td>8.6</td>
</tr>
<tr>
<td>10-19†</td>
<td>153</td>
<td>2.3</td>
<td>3.9</td>
</tr>
<tr>
<td>Mothers</td>
<td>729</td>
<td>3.5</td>
<td>18.5</td>
</tr>
<tr>
<td>Fathers</td>
<td>531</td>
<td>2.5</td>
<td>13.5</td>
</tr>
<tr>
<td>Total</td>
<td>3928</td>
<td>4.1</td>
<td>100.</td>
</tr>
</tbody>
</table>

* UR = upper respiratory; LR = lower respiratory (including illnesses with “croupy” cough); R+ = respiratory plus other including enteric, rash, etc.
† 10-19 includes a small number of nonparent adults.
The Cauqué Study

- 1964-1974 – PI was Dr. Leonardo Mata
- Purpose – Investigate the role of infection and nutrition in growth of young children in a third world setting (a Mayan village in Guatemala)
- Nevin Scrimshaw of MIT described the study as follows:
The Cauqué Study

“It is unlike any other published study in its almost daily observations of microbiological and general health status in the direct response to the cultural and environmental events to which each child is exposed from fetal life to school age.”
The Cauqué Study
The Cauqué Study
The Cauqué Study

- **Design**
  - Study took place in a Mayan village (Santa Maria Cauqué) in the highlands of Guatemala
  - 45 mothers and children followed from pregnancy of the mother to 3 months post weaning of the child
  - Weaning age ranged from 1-4 years of age

- **Data Collection**
  - IBM punch cards – over 2000 cards per child with 8 different record types
  - Data was key entered and stored on magnetic tapes and later on magnetic disks
### The Cauqué Study

- **Study Conduct**

<table>
<thead>
<tr>
<th>Table 4.4</th>
<th>Times studied</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Program of Study, Cohort Children, 1964–1972</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Growth</strong></td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td>at birth</td>
</tr>
<tr>
<td>Weight, height, circumferences of head and thorax</td>
<td>daily, first week of life</td>
</tr>
<tr>
<td></td>
<td>within 24 hours of birth</td>
</tr>
<tr>
<td></td>
<td>weekly, neonatal period</td>
</tr>
<tr>
<td></td>
<td>fortnightly, 1st to 12th month</td>
</tr>
<tr>
<td></td>
<td>every 4 weeks, 2nd and 3 years</td>
</tr>
<tr>
<td></td>
<td>every 3 months, 4th to 7th years</td>
</tr>
<tr>
<td>X-ray of hand</td>
<td>every 6 months, beginning at 6 months</td>
</tr>
<tr>
<td><strong>Diet</strong></td>
<td>weekly, birth to 3 years; if weaned, up to 3 months after complete weaning (usually in 2nd or 3rd year)</td>
</tr>
<tr>
<td>Feeding practices and nutrient intake</td>
<td></td>
</tr>
<tr>
<td><strong>Infection</strong></td>
<td>weekly, birth to 3 to 5 years</td>
</tr>
<tr>
<td>Fecal specimen</td>
<td>when certain diseases were present, birth to 5 years</td>
</tr>
<tr>
<td>Feces or other specimen</td>
<td></td>
</tr>
<tr>
<td><strong>Health</strong></td>
<td>weekly, neonatal period</td>
</tr>
<tr>
<td>Physical examination</td>
<td>fortnightly, 1st month to 5 years</td>
</tr>
<tr>
<td>Morbidity surveillance</td>
<td>weekly, birth until 3 to 5 years</td>
</tr>
<tr>
<td>Clinical examination</td>
<td>whenever symptoms were present, birth to 5 years</td>
</tr>
<tr>
<td></td>
<td>at death</td>
</tr>
</tbody>
</table>
The Cauqué Study

- **Data Processing**
  - SPHCM minicomputer used for data management and descriptive analysis
    - Required development of a general purpose database management system which we called CCSS (Conversational Computer Statistical System)
  - University of Washington mainframe computer used for statistical analyses
The Cauqué Study

• Statistical Analysis
  • Display of data for a longitudinal study was challenging and considerable effort went into providing useful displays of the data
  • Regression modeling of growth was the key analysis methodology employed
    • A fitted growth curve with 3 parameters was calculated for each child
    • Each parameter was used as the outcome variable in multiple linear regression modeling
The Cauqué Study

- Results
  - Contrary to expectations, protein intake was comparable to that of children in the US
  - Calorie intake was inadequate particularly (possibly exclusively) during periods of illness
  - Illness was the primary determinant of poor growth
Virus Watch & Cauqué Study

- Lessons Learned
  - Feasibility of data entry by the person collecting the data at the time of collection
  - Importance of data management software
  - Advantages of control over computing resources
Coronary Artery Surgery Study

- Purpose – compare coronary artery surgery to medical therapy in patients with mild stable angina and anatomically proved coronary artery disease
- Design
  - Randomized multi-center (15 centers) clinical trial
  - 780 patients randomized
  - Primary endpoint for the trial was all cause mortality
  - Registry of 24,959 consecutive angiography patients who were not randomized
Coronary Artery Surgery Study

- Study Conduct
  - Data Collection
    - 16 different paper forms
    - Estimated 1,000,000 forms total
    - Up to 4000 variables per person (approximately 200 million characters of data)
  - Data Entry
    - ‘Intelligent’ terminals at the clinical sites were used for data entry
    - Data was transmitted by modem over the phone lines to the coordinating center at 200 characters per second
    - Dedicated large minicomputer for study
Coronary Artery Surgery Study

- Statistical Analysis
  - Logistic regression
  - Cox proportional hazards modeling
Coronary Artery Surgery Study

**Codes: Part B**
- **Distal Vessel Status**
  - 0 = not diseased
  - 1 = mildly diseased
  - 2 = moderately diseased
  - 3 = markedly diseased

**Anatomic Location Codes**
- 1. Prox RCA
- 2. Mid RCA
- 3. Dist RCA
- 4. R PDA
- 5. RPLS
- 6. 1st RPL
- 7. 2nd RPL
- 8. 3rd RPL
- 9. Inf. Septal
- 10. Ac Marg
- 11. LMCA
- 12. Prox LAD
- 13. Mid LAD
- 14. Dist LAD
- 15. 1st Diag
- 16. 2nd Diag
- 17. 1st Septal
- 18. Prox CX
- 19. Dist CX
- 20. 1st Ob Marg
- 21. 2nd Ob Marg
- 22. 3rd Ob Marg
- 23. L AV
- 24. 1st LPL
- 25. 2nd LPL
- 26. 3rd LPL
- 27. LPDA

**Codes: Part C**
- **Source of vein**
  - 1 = saphenous vein (thigh)
  - 2 = saphenous vein (calf)
  - 3 = cephalic vein
  - 4 = basilic vein
  - 5 = other

- **Anastomosis (proximal)**
  - 1 = aorta
  - 2 = IMA origin
  - 3 = other vein graft
  - 4 = other

- **(distal)**
  Enter number for coronary artery segment receiving graft.

- **Endart. Check if endarterectomy performed.**
- **Flow Mean basal flow in ml/min.**
- **Quality of graft**
  - 1 = good
  - 2 = fair
  - 3 = poor

- **Anastomotic Technique**
  - 1 = running suture
  - 2 = interrupted sutures
  - 3 = mixed
**Coronary Artery Surgery Study**

**COST ANALYSIS FOR PAPER MAIL KEYPUNCH SYSTEM - 11**  
**CLINICAL SITES FOR 7-YR STUDY**

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper + NCR paper at 0.03/sheet x 1,000,000</td>
<td>$30,000</td>
</tr>
<tr>
<td>Mail costs at 0.02/sheet x 1,000,000</td>
<td>20,000</td>
</tr>
<tr>
<td>Microfilming at 1.5/sheet x 1,000,000</td>
<td>15,000</td>
</tr>
<tr>
<td>Handling-clerks at 2.5 min/sheet requires 4 persons for 7 yr at an</td>
<td>238,000</td>
</tr>
<tr>
<td>average of $8500/yr/clerk</td>
<td></td>
</tr>
<tr>
<td>Keypunch and verify at $6/hr, 10,000 characters/hr for 200,000,000</td>
<td>120,000</td>
</tr>
<tr>
<td>characters</td>
<td></td>
</tr>
<tr>
<td>Misc. - storage, office space, etc.</td>
<td>40,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$463,000</strong></td>
</tr>
</tbody>
</table>
# Coronary Artery Surgery Study

## COST ANALYSIS FOR "INTELLIGENT" TERMINAL SYSTEM - 11

### CLINICAL SITES FOR 7-YR STUDY

<table>
<thead>
<tr>
<th>Description</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 terminals</td>
<td></td>
</tr>
<tr>
<td>11 at $8,000</td>
<td>$88,000</td>
</tr>
<tr>
<td>1 at $30,000</td>
<td>30,000</td>
</tr>
<tr>
<td>1 person year of programming</td>
<td>20,000</td>
</tr>
<tr>
<td>Maintenance at average of $50/month x 11 + 1 at $122/month</td>
<td>56,400</td>
</tr>
<tr>
<td>Phone costs at 0.21/min, 5 min/wk/center 100 characters/second, average of 3000 characters/day/center transmitted</td>
<td>3,800</td>
</tr>
<tr>
<td>1 clerk for 7 yrs</td>
<td>59,500</td>
</tr>
<tr>
<td>Misc., office space, etc.</td>
<td>5,000</td>
</tr>
<tr>
<td>Bell telephone data access arrangement (7 yrs)</td>
<td>6,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$268,700</strong></td>
</tr>
</tbody>
</table>
Coronary Artery Surgery Study

Results

- “… the similarity of survival rates in the two groups of patients … lead to the conclusion that patients similar to those enrolled in this trial can safely defer bypass surgery until symptoms worsen to the point that surgical palliation is required”

- The patients who were eligible for randomization, who were not randomized, had survival curves almost identical to those from the clinical trial
Coronary Artery Surgery Study

ALL RANDOMIZED PATIENTS

PERCENT SURVIVAL

YEAR

%SURVIVAL

LEGEND

MEDICALLY ASSIGNED
SURGICALLY ASSIGNED

\( p = 0.34 \)
Coronary Artery Surgery Study
Coronary Artery Surgery Study
Medical Patients

Graph showing survival rates over years with different symbols and annotations for randomized and late surgery groups.
Coronary Artery Surgery Study
Surgical Patients
Coronary Artery Surgery Study

- Lessons learned
  - It is possible to successfully enroll and randomize patients even in the face of opposition from a segment of the medical profession.
Impossibilities' are good not to attach that label to; since, correctly understood, if we wanted to, we would be able to be able to.

Piet Hein
Coronary Artery Surgery Study

- Lessons learned
  - It is possible to successfully enroll and randomize patients even in the face of opposition from a segment of the medical profession
  - Where possible, a registry of non-randomized patients should be enrolled and followed
  - Data entry at sites using a 'computer' is preferable to centralized data entry
  - A lot of data that is collected in longitudinal studies is never analyzed
Stroke Prevention in Atrial Fibrillation

- 1987-1997 – PI’s were David Sherman, MD and Bob Hart, MD
- Purpose – compare the efficacy of aspirin or warfarin to placebo in preventing stroke in patients with atrial fibrillation
- Design of SPAF I (Group 1)
  - Multi-center (15 centers) randomized clinical trial
  - 588 patients randomized
  - Primary endpoint was time to ischemic stroke and systemic embolism
  - Evaluation of the endpoint was done by a committee that was blinded to treatment assignment
Stroke Prevention in Atrial Fibrillation

- Data Collection
  - Specially printed paper forms were used
  - Fax machines were used for data entry
  - Faxed forms were received and converted to character data on a PC at the data center

- Statistical Analyses
  - Sequential monitoring was done and evaluated by a Data Safety Monitoring Board
Stroke Prevention in Atrial Fibrillation

- Results
  - Placebo arm was stopped early due to superiority of either aspirin or warfarin
  - Estimated risk reduction of 81%!
  - The SPAF trials changed medical practice dramatically and probably resulted in 30,000 or more fewer strokes per year in the US
Stroke Prevention in Atrial Fibrillation

<table>
<thead>
<tr>
<th>Cumulative Rate of Primary Events (%)</th>
<th>Months after Randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warfarin or Aspirin</td>
<td>Placebo</td>
</tr>
</tbody>
</table>

p<0.00005
Stroke Prevention in Atrial Fibrillation

- Lessons Learned
  - Importance of DSMB and sequential monitoring
  - Study investigators should plan ahead for possible early termination of part or all of the trial
Cardiovascular Health Study
Design and Conduct

- 1989-present – NHLBI contract
- Design
  - Enrolled 5888 elderly (65-99 years of age) including over 800 African-Americans at four clinical sites
  - Extensive baseline examination
    - Echocardiogram
    - Carotid Ultrasound
    - Serum collection
    - ECG
    - Medical and physical history
    - Etc
Cardiovascular Health Study Design and Conduct

- Objectives
  1. To quantify associations of conventional and hypothesized risk factors with CHD and stroke
  2. To assess the association of indicators of subclinical disease, identified by noninvasive measures such as carotid ultrasonography and echocardiography, with incidence of CHD and stroke
  3. To quantify the association of conventional and hypothesized risk factors with subclinical disease
  4. To characterize the natural history of CHD and stroke, and identify factors associated with clinical course
  5. To describe the prevalence and distributions of risk factors, subclinical disease, and clinically diagnosed CHD and stroke
Cardiovascular Health Study
Design and Conduct

- Other measures added during study
  - MRI of the brain
  - Endothelial function
  - Retinal eye exam
  - Numerous ‘Ancillary’ studies collected other data

- Repeat examinations
  - Carotid Ultrasound - 2 times
  - Echocardiography
  - MRI of brain
  - Serum collected – 2 times

- Every 6 month follow-up contacts and careful ascertainment and classification of events
Cardiovascular Health Study
Design and Conduct

- **Data collection**
  - Local area network of PC’s at each clinical site for data entry and transmission
  - Internet used to transmit data to the coordinating center

- **Data management**
  - PC’s used for all data management and statistical analyses
  - Web site for CHS for study documentation and communication – chs-nhlbi.org
Cardiovascular Health Study Results

- More than 200 papers published with an additional 200+ in process
- Objectives of study were met
  - Much better understanding of all aspects of cardiovascular disease in the elderly
  - Extensive findings concerning subclinical cardiovascular disease
- Ancillary studies have contributed important findings (e.g. role of vascular disease in dementia)
Cardiovascular Health Study

Results

- Example of findings – Relationship of carotid wall thickness to risk of stroke and Myocardial Infarction
Cardiovascular Health Study

Results

Rate of Myocardial Infarction or Stroke per 1000 Person-Years

Quintile of Intima–Media Thickness

- Maximal CCA IMT
- Maximal ICA IMT
- Maximal CCA and ICA IMT

36.5, 36.1, 40.9
Cardiovascular Health Study

- Lessons Learned
  - Advantages and barriers to use of WWW for study management and communication
  - Importance of measurement error when studying change over time
  - Encouragement of outside collaboration enhances the output from longitudinal studies
  - Much more data is collected than is ever utilized
Present and Future

- Collaborative Health Studies Coordinating Center (CHS CC)
  - Multi-Ethnic Study of Atherosclerosis (MESA)
  - Ginkgo Evaluation of Memory (GEM) Study
- Therapeutic Development Network Coordinating Center for Cystic Fibrosis
- Azithromycin and Coronary Events Study (ACES)
Present and Future

- Trends
  - More extensive use of internet for data collection, study management and communication
  - Inclusion of more ‘high tech’ measures
  - Genetics research
  - Continued advancements in statistical methodology
  - A return of interest in studies of infectious disease
The Future?

- Lesson Learned
  - It’s impossible to accurately predict the future!
You'll conquer the present suspiciously fast if you smell of the future — and stink of the past.