Introduction to Cohort Studies

Beate Ritz MD PhD

Epi 200A

Fall 2012

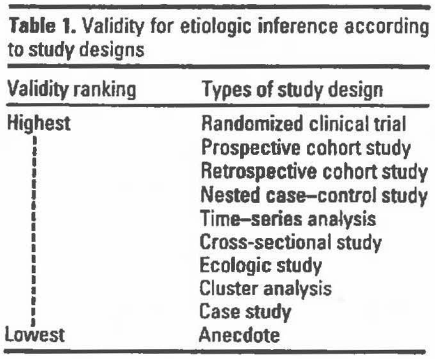


Table 1. Validity for etiologic inference according to study designs

Validity ranking	Types of study design
Highest	Randomized clinical trial
	Prospective cohort study
	Retrospective cohort study
	Nested case–control study
	Time–series analysis
	Cross-sectional study
	Ecologic study
	Cluster analysis
	Case study
Lowest	Anecdote

Source: Kunzli et al. The Semi-individual Study in Air Pollution Epidemiology: A Valid Design as Compared to Ecologic Studies. *EHP* 1997, 105 (10)

MacMahon and Pugh, 1970

Definition of cohort studies (in public health epidemiology)

- The group or groups of persons to be studied are defined in terms of **characteristics manifest prior to the appearance of the disease** under investigation
- The study group so defined are observed over a period of time to determine the **frequency of disease** among them

Cohort studies

Simplistic description

- A cause 'looking' for a disease
- (*versus* case-control study: "A disease 'looking' for a cause")

Cohort design:

Retrospective (historical) in terms of

a) timing of events *or*

b) data collection

Cohort is enumerated some time in the past and followed over *historical* time (to today)

- time of follow-up long (20-40 years), often extends across decades
- cohort can be large i.e. 10,000+ members

But, how do we:

- "reconstruct" the cohort - who belongs into the cohort?
- Obtain exposure and outcome information
  - Note: a historical cohort is often restricted to investigations of fatal disease (why!)

Cohort design:

Prospective in terms of

a) timing of events *or*

b) data collected

This design is best to be used for

- short-term (common) health outcomes; e.g. for:
  - physiological changes (blood pressure and noise)
  - acute neurotoxic effects (OP pesticides)
  - pulmonary function (cotton dust)
  - skin rashes (irritants, e.g. solvents, metals)
  - injuries
  - allergic reactions, asthma attacks
- prospective medical surveillance

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

TRIAL EXHIBIT 1467

Case No. 3:16-cv-00525-VC

Date Entered ____________

By ____________
Deputy Clerk

EXHIBIT 19-17

RITZ

Date: 9/18/2017

Reporter: Lisa Moskowitz
CSR 10816. RPR. CRR. C'

Cohort design:

Prospective *or* retrospective in terms of

a) timing of events *or*

b) data collected

The major issue we want to convey is whether disease status could have influenced exposure measurement/information (such as via recall of exposure by a diseased subject)

Note that retrospective often is considered a 'less reliable' design; thus, be clear about how you use this term

Cohort study: examples

Cohort: "Any designated group of individuals who are followed or traced over a period of time"

Historically:

- John Snow: Cholera in London (1854)
- Panium: Measles on the Faroe Islands (1846)

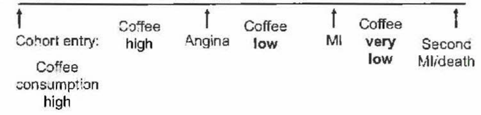
More recent

- Framingham: cardiovascular diseases (N=5,209); bi-annual exams, medical records and deaths info
- British doctors: smoking and lung cancer among British doctors (N=34,439 male British doctors in 1951; Doll)
- Perinatal collaborative study: pregnancy and child health, cerebral palsy and allied neurological defects (N=42,000 pregnant women enrolled 1959–1966 at 12 hospitals across the United States)
- Nurses Health Study: estabilsned in 1976 from female US registered nurses ages 30-55 years who responded to a mailed questionnaire that inquired about risk factors for cancer and heart disease (N=121,700)
- HIV cohorts: 1984-2005, Multicenter AIDS Cohort Study (N=4,955 homosexual men who volunteered in Baltimore, Chicago, Los Angeles, and Pittsburgh)
- EPIC study: cancer
- California Teachers Cohort (125,000 in 1995): Breast cancer
- And many more

Causal Inferences in Cohort Studies

- Since the only *sine qua non* causal criteria states that **a cause precedes its effect,** it is logical to start with the exposure and follow exposed people forward in time to study the occurrence of the health endpoint of interest
- This was hardly done *prospectively* before the Framingham cohort study (baseline 1948); too expensive, too time consuming
- A cohort study is a logical design to study determinants of the changes from not having a disease to having a disease. The study can guarantee that exposure precedes the onset of clinical diagnosis (but perhaps not the real onset of pathological changes).

Example: Does coffee drinking trigger myocardial infarction (MI)?



Cohort entry:

Coffee consumption high

Coffee **high**

Angina

Coffee **low**

MI

Coffee **very low**

Second MI/death

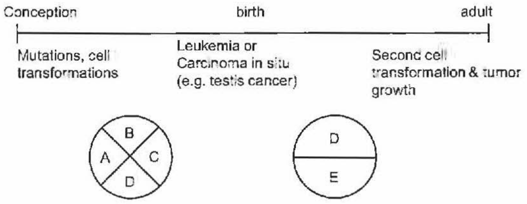
Experimental *vs.* Observational Studies:

Why not conduct a randomized trial?

Trials

- cannot obtain evidence for harmful agents (and sometimes for beneficial ones as well)
- deal by nature with (very) selected populations
- not practical for
  - rare outcomes (*Note:* we would expect only 50-200 lung or colon cancers and 16 Parkinson's cases per 100,000 person years of observation in most working age cohorts)
  - long follow-up times that allow for latency
  - effects that occur late in disease progression
- focus on one (or several) specific doses only
- expensive to conduct

Life course perspective


Conception
birth
adult
Mutations, cell transformations
Leukemia or Carcinoma in situ (e.g. testis cancer)
Second cell transformation & tumor growth
B
A
C
D
D
E

Different causal components may be operating

Note: many cohorts recruit at entry only few of those eligible (% of all eligible often not known):

- What is the impact on internal validity and external validity?

Cohort studies: recruitment

- Recruitment to the cohort may be mandatory/ automatic
  - All in public registers = mortality, births, deaths, cancer (without informed consent)
  - Occupational cohorts using employment data from occupational plants (assess exposures retrospectively from records and outcomes from registers)
- NOTE: cohorts using "primary" data (i.e. collected during/for the investigation) are usually based upon **informed consent**

Examples:

- via General Practitioner – e.g. Danish National Birth Cohort
- Letters – e.g. to members of an organization (British doctors, CA Teachers, Nurses Health Study, Harvard Alumni)
- Advertisements – e.g. people with a given disease
- Local community: ALSPAC, Framingham
- Visitors to a website
- Participants in L.A. Marathon

Cohort studies: follow-up

- Compliance to follow-up procedures
  - frequent contacts needed!
  - Are (health) benefit incentives given?
- Recording of endpoints
  - rely on diagnoses made by the health care system
  - repeated measurements necessary?
- Changes in other determinants/ covariates
  - questionnaires
  - interviews
  - measurements
- Participation is voluntary, participants are free to leave the cohort at any point in time
  - right to remove data from the study?

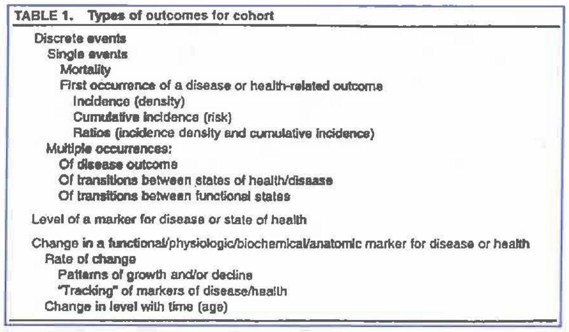


TABLE 1. Types of outcomes for cohort

- Discrete events
  - Single events
    - Mortality
    - First occurrence of a disease or health-related outcome
      - Incidence (density)
      - Cumulative incidence (risk)
      - Ratios (incidence density and cumulative incidence)
  - Multiple occurrences:
    - Of disease outcome
    - Of transitions between states of health/disease
    - Of transitions between functional states
- Level of a marker for disease or state of health
- Change in a functional/physiologic/biochemical/anatomic marker for disease or health
  - Rate of change
    - Patterns of growth and/or decline
    - "Tracking" of markers of disease/health
  - Change in level with time (age)

Source: Tager IB. Outcomes in cohort studies. *Epidemiologic Reviews* 1998, 20(1).

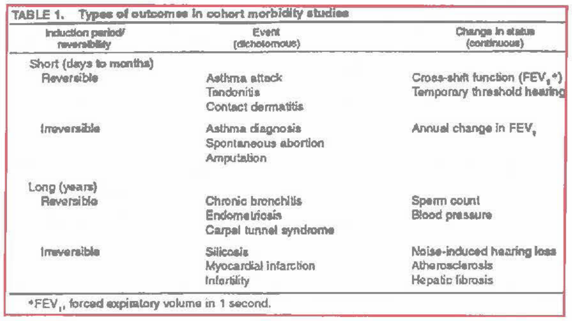


TABLE 1. Types of outcomes in cohort morbidity studies

Induction period/ reversibility	Event (dichotomous)	Change in status (continuous)
Short (days to months)		
Reversible	Asthma attack Tendonitis Contact dermatitis	Cross-shift function (FEV_1*) Temporary threshold hearing
Irreversible	Asthma diagnosis Spontaneous abortion Amputation	Annual change in FEV_1
Long (years)		
Reversible	Chronic bronchitis Endometriosis Carpal tunnel syndrome	Sperm count Blood pressure
Irreversible	Silicosis Myocardial infarction Infertility	Noise-induced hearing loss Atherosclerosis Hepatic fibrosis

*FEV_1, forced expiratory volume in 1 second.

Source: Checkoway H and Eisen EA. Developments in Occupational Cohort Studies. *Epidemiologic Reviews* 1998, 20(1).

Cohort Entry Definitions

Entry to a cohort can be defined at a fixed point in time:

- All subjects are selected at a given point (range) in time, e.g. from a registry of a type of people
  - All atomic bomb survivors in Japan on Jan 1st 1950 living in Nagasaki and Hiroshima
  - European Prospective Investigation into Cancer and Nutrition (EPIC), a multi-centre prospective cohort study in 23 study centers in ten European countries
    - E.g in Germany, recruitment was based on a random sample of subjects in targeted age range (women aged 35–65, men 40-65) from population registers between 1994 and 1998
    - participation rate was 38.5% (i.e. observed cohort is a self-selected subgroup of the underlying population)

or

- subjects enter the cohort at different points in time; *e.g.*: all inhabitants of Framingham/MA that reach a certain age

104 Research Methods in Occupational Epidemiology

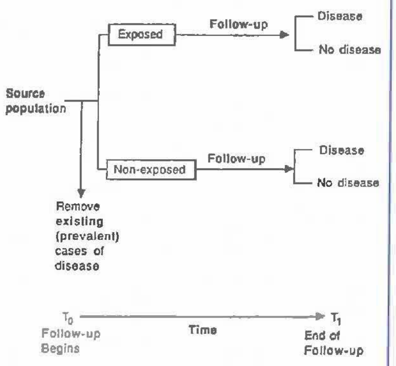


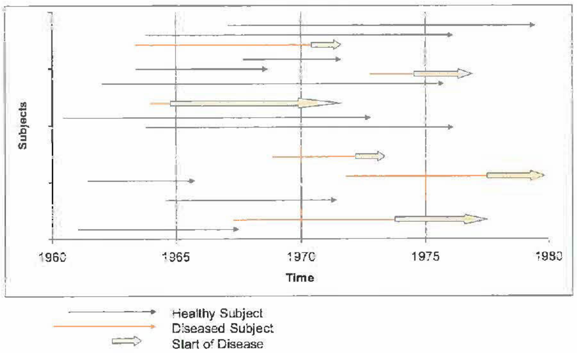
Figure 5-1. Flow diagram of cohort study design.

Cohort Exit Definitions

Subjects can be follow-up

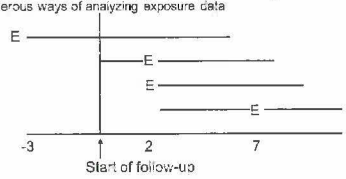
- until a fixed point in calendar time (end of study);
  - note: some subjects are observed for a shorter time i.e. due
    - incidence of the disease under investigations,
    - death,
    - migration or
    - loss of follow-up
- or as long as they are
  - employed
  - live in the city
  - have the exposure (are "right censored" when this changes) (e.g. use of a certain type of medication)

Study Design Overview: Identifying Diseased Subjects in a Population



Cohort studies: exposure assessment

- Exposure may have started at a given point in time:
  - E.g. at baseline or any other measurement point
  - and remains fixed ("ever smoker")
  - or changes over time (amount of smoking)
- Exposure can be measured as:
  - Average or cumulative exposure over time
  - exposure level at baseline
  - Note: without a prior hypothesis (or knowledge of biological mechanism) there may be numerous ways of analyzing exposure data



Person time 'lagged' by x years (after time of hire) – immortal PT

Cohort studies: exposure assessment

- Exposures can be lagged (i.e. exclude exposure during time irrelevant for the disease)
  - E.g. exposure too close to disease onset
- Exposure contrast
  - Generally we like to examine as large an exposure contrast as possible - thus, we want to establish a cohort with different exposure levels (e.g. workers in a copper-smelter compared to the general population)

- Select the non-exposed subjects as close to the **counterfactual** ideal as possible
  - Non-exposed subjects should have the same disease risk as the exposed had they not been exposed

12 Research Methods in Occupational Epidemiology

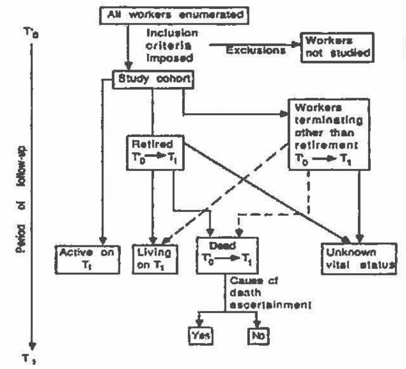

T_0
Period of follow-up
T_1
All workers enumerated
Inclusion criteria imposed
Exclusions
Workers not studied
Study cohort
Workers terminating other than retirement $T_0 \rightarrow T_1$
Retired $T_0 \rightarrow T_1$
Active on T_1
Living on T_1
Dead $T_0 \rightarrow T_1$
Unknown vital status
Cause of death ascertainment
Yes
No

Figure 5–3. Flow diagram of mortality follow-up of historical cohort.

Start of follow-up in a cohort study

- hire date or fixed time/date after hiring
- first monitoring date (e.g. radiation monitoring, blood lead monitoring)
- fixed date (such as Jan 1970)

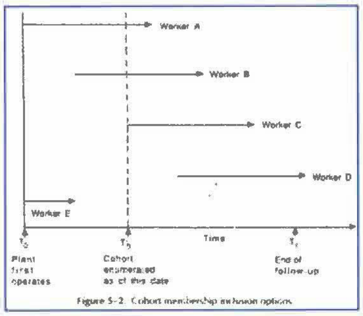


Figure 5–2. Cohort membership inclusion options.

End of follow-up in a cohort study

- end of follow-up for the cohort reached
- death or incidence from outcome of interest
- death from competing causes
- last known date alive (after that we call them 'lost to follow up')

Or

- should we assume a worker is alive if no information is found that indicates that the subject died (and thus continues to add person-time)?

Pearce et al.

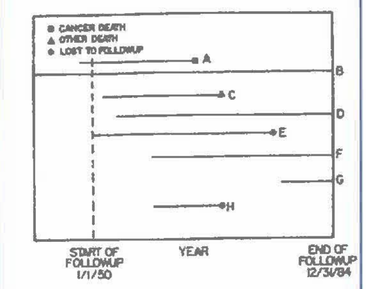


Fig. 1. Follow-up over a 34 year-period of a hypothetical cohort.

12 Research Methods in Occupational Epidemiology

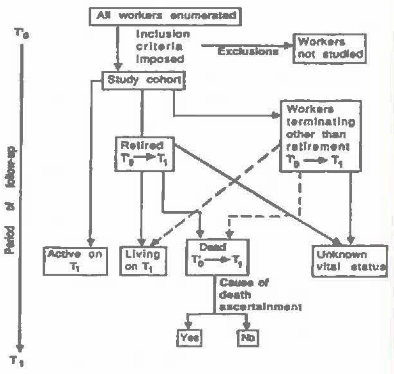

T'0
All workers enumerated
Inclusion criteria imposed
Exclusions
Workers not studied
Study cohort
Workers terminating other than retirement $T_0 \rightarrow T_1$
Retired $T_0 \rightarrow T_1$
Period of follow-up
Active on T_1
Living on T_1
Dead $T'_0 \rightarrow T_1$
Unknown vital status
Cause of death ascertainment
Yes
No
T_1

Figure 5–3. Flow diagram of mortality follow-up of historical cohort.

Summary: Cohort Studies

- Generally most accepted in scientific community
- Include the entire available study population
- Most similar to standard experimental strategies
  - determine (rather than apply) a toxin or preventative agent among subjects disease-free at baseline
  - follow-up subjects over time
  - observe adverse or positive health effects in exposed and non-exposed subjects

The goal is to estimate the risk of (various or one) disease/s among the exposed subjects relative to the background risk experienced by "comparable" unexposed persons:

- comparable refers to the "exchangeability assumption" or "counterfactual"
  - *what would have happened to this group of exposed subjects if they had NOT been exposed?*

Summary: Cohort Studies

- Select non-exposed as close to the counterfactual ideal as possible:
  - Non-exposed should have the same disease risk as the exposed had they not been exposed
- Recruitment to the cohort
  - based upon informed consent if primary data are collected
  - Without informed consent if all are followed in public registers = mortality, births, deaths
- Historical cohorts: e.g. use existing data but need not be 'retrospective'

Advantages of the cohort method

- In principle, can provide a **complete description of experience of cohort members** subsequent to exposure, including rates of progression to and staging of disease, and natural history of disease
- Allows study of **multiple potential effects** of a given exposure, thereby obtaining information on potential benefits as well as risks
- Allows for the **calculation of rates** of disease in exposed and unexposed individuals and time to event
- Permits **flexibility in choosing variables** to be systematically recorded
- Allows for **thorough quality control in measurement** of study variables (not in historical cohort studies though)

Disadvantages of the cohort method

- Large numbers of subjects required (thus, low feasibility to study rare diseases)
- Relatively expensive to conduct
- Potentially long duration for follow-up necessary
- Exposures may change, making findings irrelevant unless the exposure assessment is adapted
- Maintaining follow-up may be difficult
- The cohort is generally not representative of the general population

AGRICULTURAL HEALTH STUDY

Example: The Agricultural Health Study Cohort (AHS)

- Collaborative effort to study the effects of pesticide exposures among farmers
  - National Cancer Society (NCI)
  - National Institute of Environmental Health Sciences (NIEHS)
  - U.S. Environmental Protection Agency (EPA)

http://aghealth.nci.nih.gov/

The AHS Cohort study:
Retro- and prospective data collection

- Phase I, initial cohort recruitment, 1994-1997:
- 89,658
  - private pesticide applicators, and
  - spouses of private applicators, and
  - commercial pesticide applicators
- Recruited at Iowa and North Carolina state pesticide applicator licensing facilities
- Each pesticide applicator asked to complete a 21-page enrollment questionnaire

a. Demographic data
b. Pesticides used (50 pesticides), other pesticide-related questions
c. Lifestyle (i.e., smoking, alcohol, vegetable, and fruit consumption)
d. Brief medical history
e. Family history of cancer, kidney failure, diabetes, and heart disease
f. Farm exposures other than pesticides (not in commercial pesticide applicator version)
g. Personal identifiers, spouse identifiers, children identifiers

Farmer applicators completing the enrollment questionnaire are given three take-home questionnaires (scanable) for

- the applicator (licensing exam taker)
- spouse, and
- female and family health questionnaires

The AHS Cohort

Take Home Questionnaires:
Farmer Applicator/Commercial Applicator

a. Farm exposures (comprehensive)
b. Pesticide use information (i.e., methods of application, additional pesticides used)
c. Work practices used currently versus those used 10 years ago
d. Other occupational exposures
e. Leisure and work physical activity, physical attributes (e.g., height, weight, eye color, skin pigmentation category)
f. Dietary and cooking practices
g. Medical history (comprehensive)
f. Personal identifiers

The AHS Cohort

- Cancer and non-cancer outcomes
  - Linkage with
    - » cancer registries
    - » vital statistics
    - » United States Renal Data System (USRDS)
  - Exposure data collection
    - » Baseline questionnaire at licensing exam
    - At follow-up
    - » telephone interviews (CATI)
    - » food frequency questionnaire and
    - » cheek cell collection

- Phase II: follow-up in 1999-2003
- Phase III: follow-up in 2004-2008

The AHS Cohort

1. Cohort studies
   - □ All cause and cancer mortality
   - □ cancer incidence
2. Cross-sectional studies:
   - □ Using questionnaire data, functional measures, biomarkers, and GIS
   - □ E.g. cross sectional immunology study of atrazine applicators/corn farmers in Iowa
3. Nested case-control studies
   - □ High pesticide exposure events
   - □ Parkinson's disease study
4. Exposure assessment and validation studies

The AHS Cohort

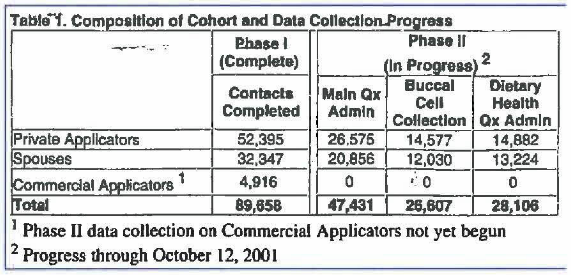


Table 1. Composition of Cohort and Data Collection Progress

	Phase I (Complete)	Phase II (In Progress)[2]		
	Contacts Completed	Main Qx Admin	Buccal Cell Collection	Dietary Health Qx Admin
Private Applicators	52,395	26,575	14,577	14,882
Spouses	32,347	20,856	12,030	13,224
Commercial Applicators[1]	4,916	0	0	0
Total	**89,658**	**47,431**	**26,607**	**28,106**

[1] Phase II data collection on Commercial Applicators not yet begun

[2] Progress through October 12, 2001

The AHS Cohort

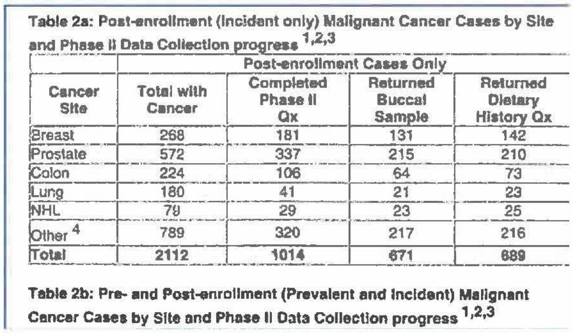


Table 2a: Post-enrollment (incident only) Malignant Cancer Cases by Site and Phase II Data Collection progress[1,2,3]

	Post-enrollment Cases Only			
Cancer Site	Total with Cancer	Completed Phase II Qx	Returned Buccal Sample	Returned Dietary History Qx
Breast	268	181	131	142
Prostate	572	337	215	210
Colon	224	106	64	73
Lung	180	41	21	23
NHL	79	29	23	25
Other[4]	789	320	217	216
Total	2112	1014	671	689

Table 2b: Pre- and Post-enrollment (Prevalent and Incident) Malignant Cancer Cases by Site and Phase II Data Collection progress[1,2,3]

Ag-Health study topics

Cancer mortality and incidence in Applicators and Spouses

Pesticide Exposure Assessment, Applicators, Spouses and Children – questionnaires

Pesticide Exposure Assessment - Field Studies – Acute exposures

Biologic and Functional Effects of Chronic Pesticide Exposure

Biomarkers and Molecular Genetics

Injury

Lifestyle and Diet

Non-pesticide Exposures, Exposure to Animals

Respiratory Disease and Function

Neurological Disease and Function

Reproductive Health, Child and Adolescent Health

Autoimmune Disease and Immune Function

Other Non-cancer Chronic Disease

Pooling of cohorts

Advantages:

- Can study rare outcomes
- Conduct subgroup analyses for effect measure modifiers (e.g. sex, race etc)
- Wide geographic distribution allows spread of exposures
- Availability of prospective data; stored serum blood samples can be analyzed by same lab

Disadvantages

- Usually no common data elements, i.e. diverse data collection methods need to be reconciled
- Some variables may not have been collected at all; how to handle missing data?

Vid D and type 2 diabetes: meta-analysis

Table 1 General characteristics of the prospective studies included in the review

Lead author	Publication date	Study name	Location	Time period	Population source	Baseline age (years)	Male (%)	Follow up (years)	No. of participants	No. of events
(1) Type 2 diabetes										
Gagnon	2011	AusDiab	Australia	1999–2000	Population register	≥ 25	45	5	5200	199
Liu	2010	FHS Offspring	United States	1987–1998	Population register	NR	46	7	2956	135
Knekt	2008	FMCHES	Finland	1973–1976	Population register	40–74	49	22	3327	230
Husemoen	2012	Inter99	Denmark	1999–2001	Population register	30–65	48	5	5726	141
Kim	2009	JPHC	Japan	1990–1993	Population register	40–69	43	5	59796	1114
Knekt	2008	M-FHS	Finland	1978–1980	Population register	40–69	47	17	4176	188
Thorand	2011	MONICA/KORA Augsburg	Germany	1984–1995	Population register	35–74	52	11	1883	416
Pittas (1)	2010	Nurses Health Study	United States	1989–1990	Employee register	43–70	0	14	1167	608
Pittas (2)	2006	Nurses Health Study	United States	1980	Employee register	30–55	0	20	83 779	4843
Kayaniyil	2011	PROMISE	Canada	2004–2006	Population register	≥ 30	NR	3	489	30
Deleskog	2012	SDPP	Sweden	1992–1998	Population register	35–56	60	8–10	1000	134
Grimnes	2010	Tromso study	Norway	1994–1995	Population register	≥ 25	NR	11	6119	247
Robinson	2011	WHI	United States	1993–1998	Trial register	50–79	0	7	5140	317
Liu	2005	WHS	United States	NR	Population register	45–75	0	9	10 066	805
Sub-total									190 826	8887
(2) Other metabolic outcomes										
Gagnon	2012	AusDiab	Australia	1999–2000	Population register	≥ 25	42	5	4164	523
Fung	2012	CARDIA	United States	1985–1986	Healthcare register	18–30	55	20	4727	583
Forouhi	2008	MRCEPS	United Kingdom	1990–1992	Population register	40–69	41	10	524	541
Liu	2005	WHS	United States	NR	Population register	≥ 45	0	9	10 066	5080
Sub-total									15 481	6580

AusDiab, Australian Diabetes, Obesity and Lifestyle Study; CARDIA, Coronary Artery Risk Development in Young Adults Study; DHC, FHS, Framingham Heart Study; FMCHES, Finnish Mobile Clinic Health Examination Survey; JPHC, Japan Public Health Centre-based Prospective Study; M-FHS, Mini-Finland Health Survey; MONICA/KORA Augsburg, Monitoring of Trends and Determinants in Cardiovascular Disease Cooperative Health Research in the Region of Augsburg; MRCEPS, Medical Research Council Ely Prospective Study; MS, Metabolic Syndrome; NR, PROMISE, PROspective Metabolism and ISlet cell Evaluation Cohort Study; SDPP, Stockholm Diabetes Prevention Program; WHI, Women's Health Initiative; WHS, Women's Health Study

*Also assessed insulin resistance outcomes.

†This is the number of incident cases of Type 2 diabetes. MS was derived as a continuous risk z-score in this study.

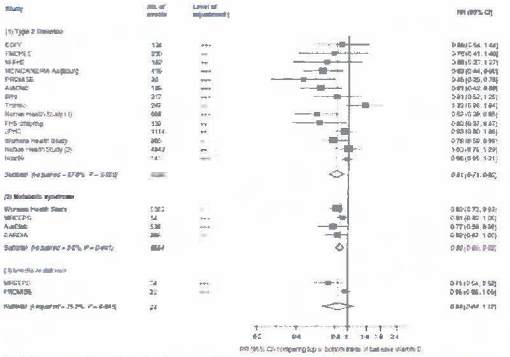


Fig. 2. Associations of vitamin D with Type 2 diabetes and other metabolic outcomes. Study acronyms and references are provided in Table 1. The summary estimate presented was calculated using a random effects model; using a fixed effects model was 0·96 (95% CI 0·94, 0·99) for diabetes, 0·86 (95% CI 0·80, 0·92) for metabolic syndrome (MS), and 0·92 (95% CI 0·83, 1·01) for insulin resistance outcomes. †Level of adjustment: +, adjusted for age and sex; + +, adjusted for diabetes/MS risk factors; + + +, adjusted for diabetes/MS risk factors plus seasonality/latitude. These studies quantified insulin resistance using homeostasis model of assessment–insulin resistance, a continuous variable.

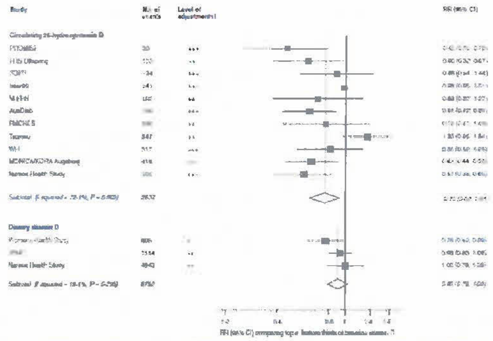


Fig. 3. Association between vitamin D and Type 2 diabetes, measured as circulating blood and dietary exposures. Study acronyms and references are provided in Table 1. The summary estimate presented was calculated using a random effects model, using a fixed effects model was 0·97 (95% CI 0·94, 1·00) for 25-hydroxyvitamin D and 0·91 (95% CI 0·81, 1·02) for dietary vitamin D intake. †Level of adjustment: +, adjusted for age and sex; + +, adjusted for diabetes/metabolic syndrome (MS) risk factors; + + +, adjusted for diabetes/MS risk factors plus seasonality/latitude.

Person time

Incidence Proportion: A/N   A= case number N=initial population size

Person-time instead of persons:

A/T observed rate [A= observed cases and T= person-time units in study group]

Poisson model

Pr(A=a)= exp(-I*T)(I*T)^a/a!

I = the rate parameter (average rate we would observe if we repeated the study over and over under the same conditions with the same amount of person-time T observed each time(i.e. end the follow-up when we reach T)

Note: Under the Poisson model A/T is the MLE estimator of I

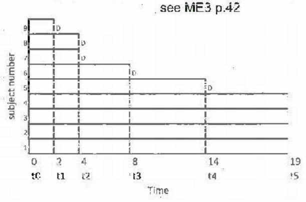
Immortal person time

The study has a criterion for a minimum of time before a subject is eligible to be in the study:

E.g. in occupational cohort studies when workers are required to have worked for a minimum of x-years. All workers who did not work for this length of time are automatically not enrolled in this cohort and all of those who are could not be censored prior to 2 years i.e. could not have died if included in the cohort.

This time should not be used to calculate person-time for those included in the cohort

Figure 3-4: Example of a small closed population with end of follow-up at 19 years
see ME3 p.42

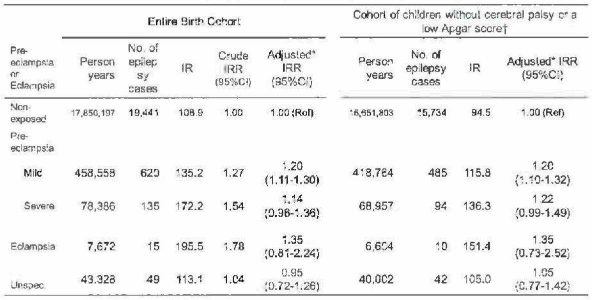


108 PY total

	Start	Outcome event times (tk)				End
	0	2	4	8	14	19
Index (k)	0	1	2	3	4	5
No. of outcome events (Ak)	0	1	2	1	1	0
No. at risk (Nk)	9	9	8	6	5	4
Prop. Surviving (Sk)		8/9	6/8	5/6	4/5	4/4
Length of interval (Δtk)		2	2	4	6	5
Person time(NkΔtk)		18	16	24	30	20
Incidence rate (Ik)		1/18	2/16	1/24	1/30	0/20

EXAMPLE:

Incidence rate ratios (IRR) for epilepsy among children exposed to pre-eclampsia or eclampsia

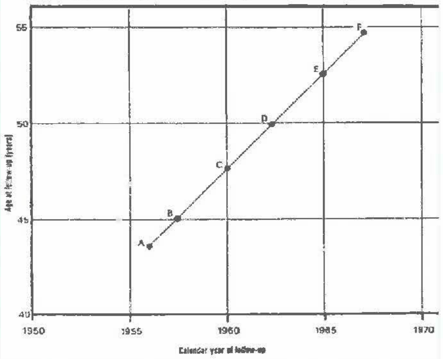


	Entire Birth Cohort					Cohort of children without cerebral palsy or a low Apgar score†			
Pre-eclampsia or Eclampsia	Person years	No. of epilepsy cases	IR	Crude IRR (95%CI)	Adjusted* IRR (95%CI)	Person years	No. of epilepsy cases	IR	Adjusted* IRR (95%CI)
Non-exposed	17,850,197	19,441	108.9	1.00	1.00 (Ref)	16,651,803	15,734	94.5	1.00 (Ref)
Pre-eclampsia									
Mild	458,558	620	135.2	1.27	1.20 (1.11-1.30)	418,764	485	115.8	1.20 (1.10-1.32)
Severe	78,386	135	172.2	1.54	1.14 (0.96-1.36)	68,957	94	136.3	1.22 (0.99-1.49)
Eclampsia	7,672	15	195.5	1.78	1.35 (0.81-2.24)	6,604	10	151.4	1.35 (0.73-2.52)
Unspec.	43.328	49	113.1	1.04	0.95 (0.72-1.26)	40,002	42	105.0	1.05 (0.77-1.42)

IR: incidence rate /100,000 person years

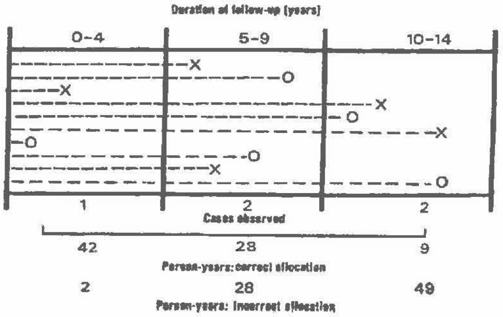
50 BRESLOW AND DAY

Fig. 2.1 Schema showing the follow-up of one person in a cohort study



BRESLOW AND DAY

Fig. 3.1 Schematic diagram illustrating proper and improper methods of allocation of person–years. ×, death from cause of interest; O, withdrawal


Duration of follow-up (years)
0–4
5–9
10–14
X
O
X
X
O
X
O
O
X
O
1
2
2
Cases observed
42
28
9
Person-years: correct allocation
2
28
49
Person-years: incorrect allocation

Person-time calculations

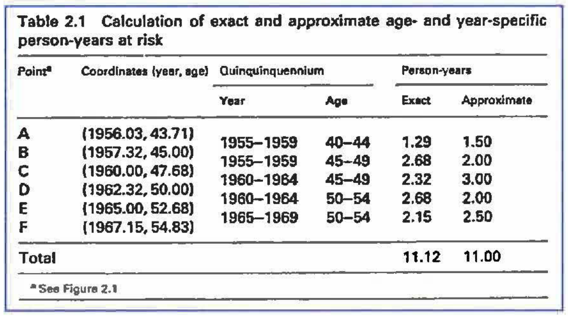


Table 2.1 Calculation of exact and approximate age- and year-specific person-years at risk

Point[a]	Coordinates (year, age)	Quinquinquennium		Person-years	
		Year	Age	Exact	Approximate
A	(1956.03, 43.71)	1955–1959	40–44	1.29	1.50
B	(1957.32, 45.00)	1955–1959	45–49	2.68	2.00
C	(1960.00, 47.68)	1960–1964	45–49	2.32	3.00
D	(1962.32, 50.00)	1960–1964	50–54	2.68	2.00
E	(1965.00, 52.68)	1965–1969	50–54	2.15	2.50
F	(1967.15, 54.83)				
Total				11.12	11.00

[a] See Figure 2.1

Incorrect *vs.* correct person-time calculations

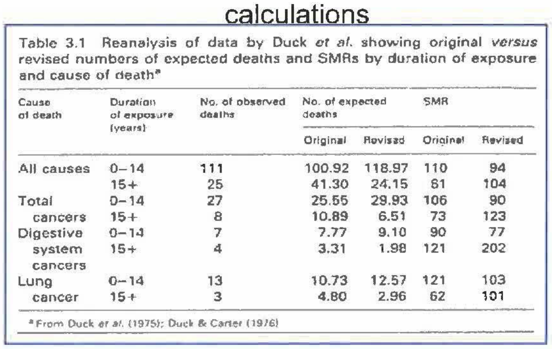
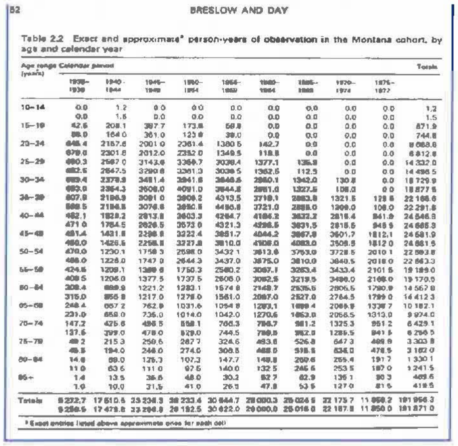


Table 3.1 Reanalysis of data by Duck *et al.* showing original *versus* revised numbers of expected deaths and SMRs by duration of exposure and cause of death[a]

Cause of death	Duration of exposure (years)	No. of observed deaths	No. of expected deaths		SMR	
			Original	Revised	Original	Revised
All causes	0–14	111	100.92	118.97	110	94
	15+	25	41.30	24.15	61	104
Total cancers	0–14	27	25.55	29.93	106	90
	15+	8	10.89	6.51	73	123
Digestive system cancers	0–14	7	7.77	9.10	90	77
	15+	4	3.31	1.98	121	202
Lung cancer	0–14	13	10.73	12.57	121	103
	15+	3	4.80	2.96	62	101

[a] From Duck *et al.* (1975); Duck & Carter (1976)

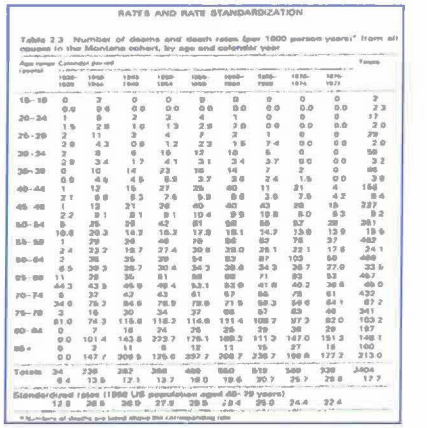


52 BRESLOW AND DAY

Table 2.2 Exact and approximate[a] person-years of observation in the Montana cohort, by age and calendar year

Age range (years)	Calendar period 1938–1939	1940–1944	1945–1949	1950–1954	1955–1959	1960–1964	1965–1969	1970–1974	1975–1977	Totals
10–14	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2
	0.0	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5
15–19	42.6	208.1	387.7	173.8	59.8	0.0	0.0	0.0	0.0	871.9
	58.0	164.0	361.0	123.0	39.0	0.0	0.0	0.0	0.0	744.8
20–24	846.4	2187.6	2001.0	2381.4	1380.5	142.7	0.0	0.0	0.0	8 939.6
	679.0	2301.8	2012.0	2352.0	1349.5	118.8	0.0	0.0	0.0	8 812.8
25–29	690.3	2587.0	3143.8	3360.7	3038.4	1377.1	135.8	0.0	0.0	14 332.0
	682.5	2647.5	3290.8	3361.3	3038.5	1362.5	112.5	0.0	0.0	14 495.5
30–34	659.4	2379.9	3481.4	3941.8	3848.5	2860.1	1342.0	130.8	0.0	18 729.9
	653.0	2364.3	3606.0	4091.0	3844.8	2851.0	1327.5	108.0	0.0	18 877.5
35–39	607.8	2186.9	3091.0	3909.2	4313.5	3718.1	2883.8	1321.5	128.6	22 188.6
	598.5	2184.8	3076.8	3990.8	4456.8	3721.0	2888.0	1309.0	108.0	22 291.8
40–44	482.1	1828.2	2813.8	3603.3	4264.7	4186.2	3632.2	2818.4	841.9	24 546.8
	471.0	1784.5	2826.5	3573.0	4321.3	4296.5	3631.5	2815.5	848.5	24 668.3
45–49	481.4	1431.8	2298.8	3222.4	3861.7	4044.2	3867.8	3601.7	1812.1	24 581.9
	480.0	1426.5	2256.8	3227.8	3810.0	4108.0	4083.0	3508.5	1812.0	24 681.5
50–54	470.0	1230.1	1758.3	2598.0	3432.1	3813.6	3763.0	3728.5	2010.1	22 803.8
	488.0	1228.0	1747.0	2644.3	3437.0	3875.0	3810.0	3840.5	2018.0	22 863.3
55–59	424.8	1209.1	1389.6	1750.3	2580.2	3067.1	3263.4	3433.4	2101.5	19 189.0
	408.5	1206.0	1377.5	1737.5	2606.0	3082.5	3219.5	3498.0	2108.0	19 170.5
60–64	308.4	889.9	1221.2	1283.1	1574.8	2148.7	2535.6	2806.5	1790.9	14 567.8
	315.0	856.8	1217.0	1278.0	1561.0	2087.0	2527.0	2764.5	1799.0	14 412.3
65–69	248.4	667.2	762.9	1031.6	1054.8	1283.1	1899.4	2068.8	1336.7	10 182.1
	231.0	658.0	736.0	1014.0	1042.0	1270.6	1863.0	2056.5	1313.0	9 874.0
70–74	147.2	425.6	496.5	558.1	760.3	786.7	981.2	1325.3	951.2	6 429.1
	137.6	399.0	478.0	529.0	744.5	780.5	982.0	1285.5	941.5	6 266.5
75–79	48.2	215.3	250.6	287.7	324.6	483.6	526.8	647.3	486.8	3 303.8
	45.5	194.0	248.0	274.0	308.5	488.0	518.5	634.0	476.5	3 182.0
80–84	14.6	89.0	125.3	107.3	147.7	148.8	200.6	265.4	191.7	1 330.1
	11.0	63.6	111.0	97.5	140.0	132.5	246.5	253.5	187.0	1 241.5
85+	1.4	13.5	36.8	48.0	30.2	62.7	62.9	136.1	90.3	463.6
	1.0	10.0	31.5	41.0	26.5	47.8	53.5	127.0	81.5	419.5
Totals	8 272.7	17 610.5	23 236.3	28 233.4	30 644.7	28 000.3	26 024.5	22 175.7	11 858.2	191 956.3
	8 258.5	17 479.8	23 298.8	28 182.5	30 622.0	28 000.0	26 016.0	22 187.8	11 850.0	191 871.0

[a] Exact entries listed above approximate ones for each cell



RATES AND RATE STANDARDIZATION

Table 2.3 Number of deaths and death rates (per 1000 person-years)[a] from all causes in the Montana cohort, by age and calendar year

Age range (years)	Calendar period 1938–1939	1940–1944	1945–1949	1950–1954	1955–1959	1960–1964	1965–1969	1970–1974	1975–1977	Totals
15–19	0	2	0	0	0	0	0	0	0	2
	0.0	9.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.3
20–24	1	6	2	3	4	1	0	0	0	17
	1.5	2.6	1.0	1.3	2.9	7.0	0.0	0.0	0.0	2.0
25–29	2	11	2	4	7	2	1	0	0	29
	2.9	4.3	0.6	1.2	2.3	1.5	7.4	0.0	0.0	2.0
30–34	2	8	6	16	12	10	5	0	0	59
	3.1	3.4	1.7	4.1	3.1	3.4	3.7	0.0	0.0	3.2
35–39	0	10	14	23	15	14	7	2	0	85
	0.0	4.6	4.5	5.8	3.7	3.8	2.4	1.5	0.0	3.8
40–44	1	12	15	27	25	40	11	21	4	156
	2.1	6.7	5.3	7.6	5.8	9.3	3.0	7.5	4.7	6.3
45–49	1	13	21	29	40	40	43	28	15	227
	2.2	9.1	9.1	9.1	10.4	9.9	10.8	8.0	8.3	9.2
50–54	5	25	26	42	61	68	56	57	26	361
	10.6	20.3	14.7	16.2	17.6	18.1	14.7	15.0	12.9	15.6
55–59	1	29	26	49	79	98	82	76	37	482
	2.4	23.7	18.7	27.4	30.6	28.0	25.1	22.1	17.6	24.1
60–64	2	35	35	39	54	83	87	103	50	488
	6.5	39.3	28.7	30.4	34.3	38.0	34.3	36.7	27.0	33.5
65–69	11	29	35	51	55	68	71	83	53	467
	44.3	43.6	45.9	49.4	53.1	53.9	41.8	40.3	39.8	46.0
70–74	5	32	42	43	61	57	66	75	61	432
	34.0	75.2	84.6	78.9	79.8	71.8	68.3	56.6	64.1	67.2
75–79	3	16	30	34	37	65	57	63	40	341
	61.0	74.3	115.5	118.2	114.0	131.4	108.7	97.3	82.0	103.2
80–84	0	7	18	24	26	25	29	38	29	197
	0.0	101.4	143.6	223.7	176.1	188.3	111.3	147.0	151.3	148.1
85+	0	2	11	6	12	11	15	27	16	100
	0.0	147.7	309.9	125.0	397.7	209.7	238.7	199.8	177.2	213.0
Totals	34	236	282	388	480	550	519	560	339	3404
	4.1	13.5	12.1	13.7	16.0	19.6	20.7	25.7	28.6	17.7
Standardized rates (1960 US population aged 45–79 years)										
	12.8	26.5	26.9	27.9	29.5	23.4	26.0	24.4	22.4	

[a] Numbers of deaths are listed above the corresponding rate

Role of Statistical Modeling

Construction of a probability model that explicitly recognizes

- the role of chance mechanism in producing some variation in the rates;
- i.e. observed rates are regarded as just one of the many possible realizations of an underlying random process.

Parameters in the model describe systematic effects of

- exposure of interest
- confounding variables such as age, period, length of follow-up etc.

Estimates of these parameters, obtained during the process of fitting the model, serve as summary statistics analogous to SMR or MH estimates of relative risk.

Role of Statistical Modeling

Advantage of model fitting over standardization:

- facilitates simultaneous consideration of several different exposure variables at risk
- estimates of relative risk obtained by model fitting generally have greater numerical stability than those computed from standardized rates.

Disadvantage of model fitting:

- parametric specification of the model due to statistical rather than biological criteria. Note: epidemiologic data are rarely extensive enough to allow to discriminate between closely related models (according to model fit criteria).

Risk set approach in a cohort study

- each subject that enters the cohort at some *entry time* is *at risk*
- each subject exits the study either as a ***failure*** i.e. contracting or dying of the disease of interest or is *censored*, i.e. is alive at the end of study, is lost to follow-up or does not contract the disease
- associated with each subject is a covariate history – fixed or time-dependent --, including factors that are known or believed to be related to the rate of the disease of interest
- At each failure a *risk set* is formed of the size *m* that included the *case* (failure at that failure time) and all *controls*, i.e. any other cohort member who is at *risk* at the failure time.

Note: The approach that organizes the cohort data by risk sets leads to data which looks just like a matched case-control study and hence we can use the conditional logistic likelihood for the analysis

also note: the risk sets are not independent, i.e. subjects can be sampled as controls in multiple risk sets and failures can serve as controls in risk sets prior to their failure times.

Risk set approach in a cohort study

Confounder control can be achieved by either

- Modeling the effect of the confounder
- Restricting each risk set to those who have similar (or the same) confounder values (=matching).

Note: if the matching factors are categorical this approach corresponds to stratification in the Cox model

Sampling from Risk Sets

- Risk set sampling designs are intrinsically related to semiparametric estimation methods for parameters in the Cox proportional hazards model used in the analysis of full cohort data.

- A sampled risk set of size *m* is a subset of the risk set that contains
  - the case and *m-1* sampled controls
  - e.g. 1:1 simple nested case-control sampling: each risk set consists of the case and one control randomly sampled from all the controls in the risk set.

  note: one can use the '*(m-1)/m*' relative efficiency rule for control sampling versus full cohort analysis for testing associations between single exposures and diseases (Breslow and Patton , 1979)
  - Thus, we have for 1 case and 4 controls (or 4/5=0.8 or 80% efficiency but then for one case and 5 controls 5/6=0.83 or 83% power, and for 9/10=0.90 or 90% power, thus, we need to add 4 controls to gain10% efficiency, i.e double your efforts to increase efficiency only slightly; it gets worse after that add another 10 controls and you get 19/20=0.95 only 5% efficiency added