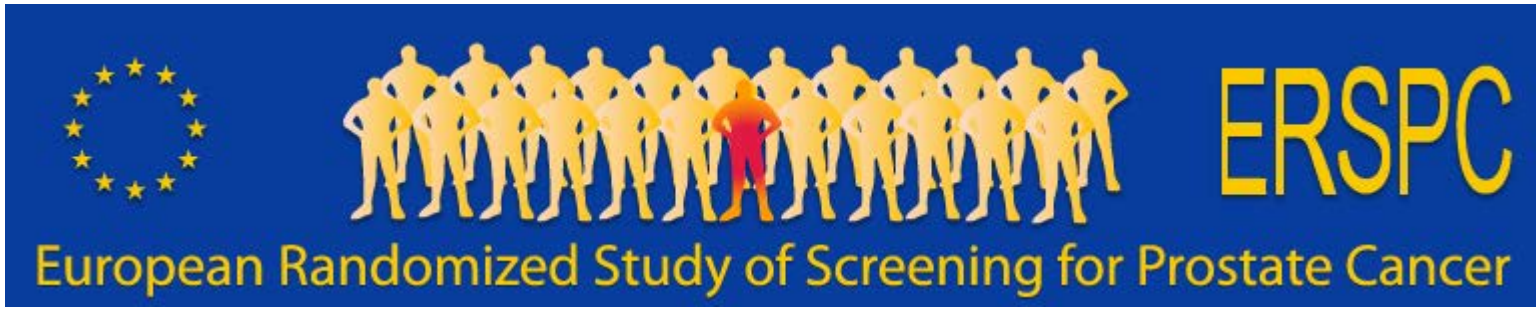
Estimating the benefits of prostate screening.

S.D. Walter
Clinical Epidemiology and Biostatistics,
McMaster University, Hamilton,
Ontario, Canada

Melbourne, April 2017

Motivation for this analysis



ERSPC – 21% reduction in prostate cancer mortality

PLCO trial – small increase in prostate cancer mortality

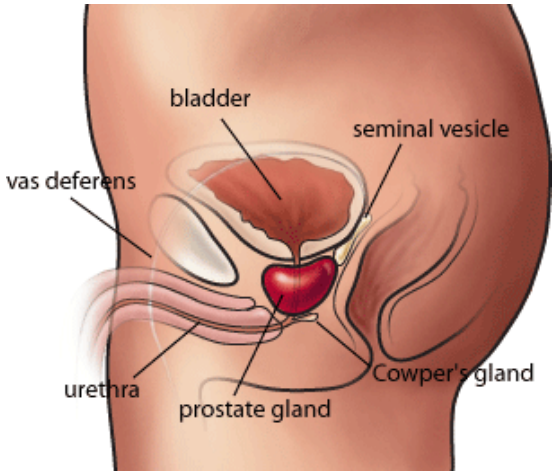
Accurate classification of the underlying cause of death is crucial for the evaluation of PC screening, but it may be unreliable.

Prerequisites for screening (WHO)

- Target disease
  - Major public health problem
  - Natural course and prognosis well known
  - Early disease can be effectively treated
- Screening test
  - Validity: High sensitivity and specificity
    - Few false positives and false negatives
  - Minimal harms, acceptable to population

Prostate cancer

- Most common cancer in men
- Globally 14 million cases, 8 million deaths (2012)
  - 1st in incidence and 2nd / 3rd in cancer deaths among men
  - Similar patterns regionally

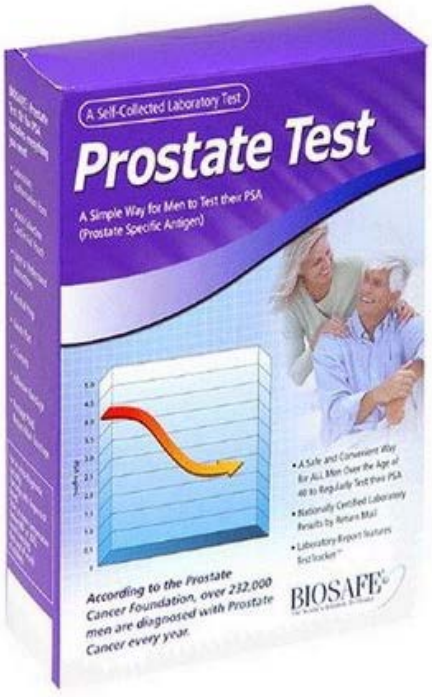


Natural course

- Common chance finding at autopsy
  - More than 50% of men aged 60+ years
- Incidence >> mortality (~1:7)
  - Many men die with the disease, but not from it
- Natural course not well understood
  - How to distinguish indolent, non-progressive disease?
- Who would benefit from treatment?
  - Objective of screening is **not** to detect all prevalent cases, but only those that require treatment

PSA as screening test

- One of the best cancer biomarkers
- Sensitivity: 73-95% (4 ng/ml)
- Specificity ~85-90%
- AUC 0.6-0.7
- Positive predictive value 20-25%



PC mortality as a study endpoint…

- About 90% five-year survival in W. Europe and North America
- Most men diagnosed with PC die from other causes
- Adjudication of the underlying cause of death is uncertain
- Inaccurate adjudication might affect study results

Outline

1. We analysed the variation in ERSPC adjudicated causes of death.

2. We used data from individual adjudicators, and the committee consensus on each case.

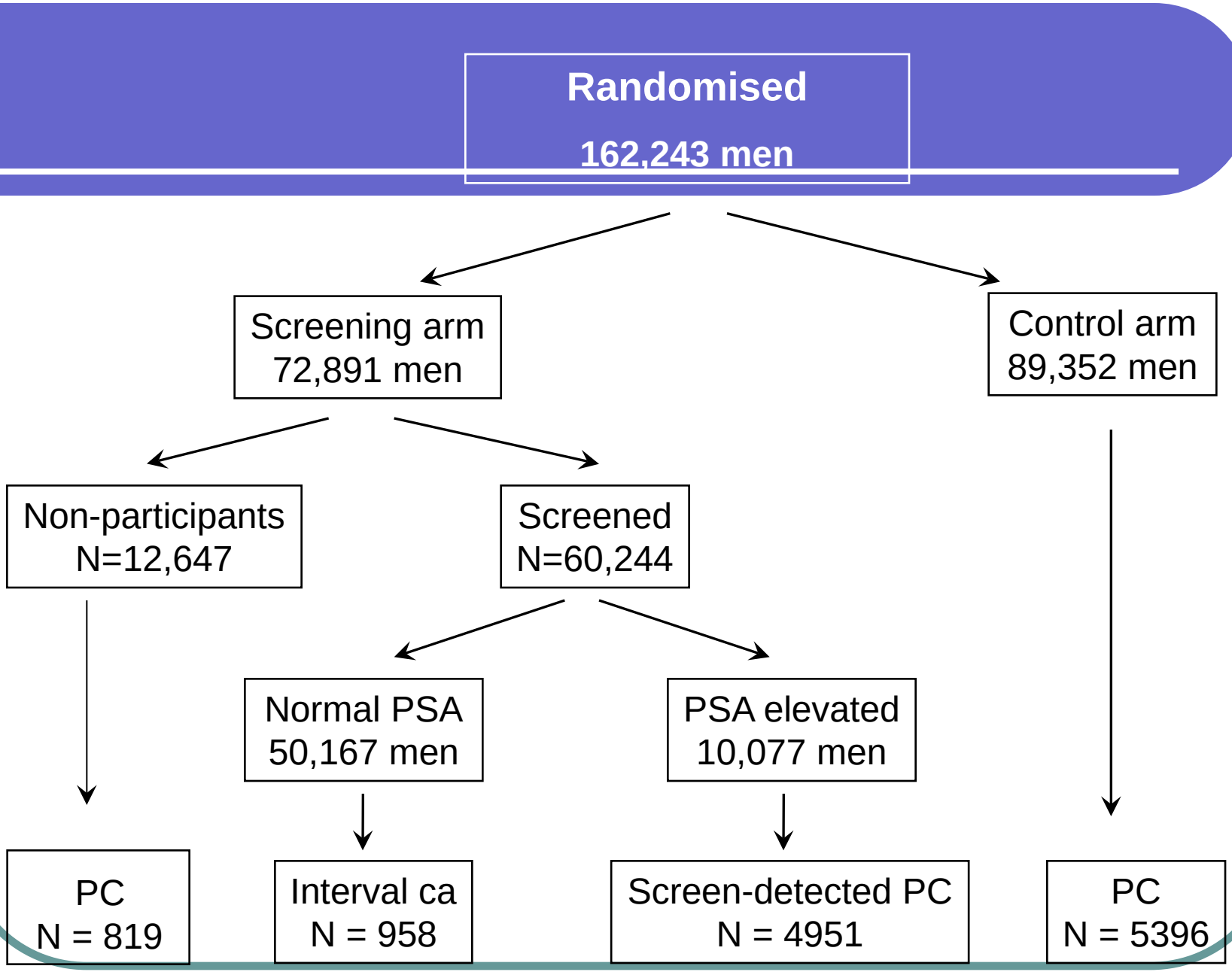
3. Latent class models (LCMs) were formulated to:

- **assess the accuracy of individual adjudicators**
- **determine if they varied significantly in accuracy,**
- **assess if accuracy might have differed between study arms.**

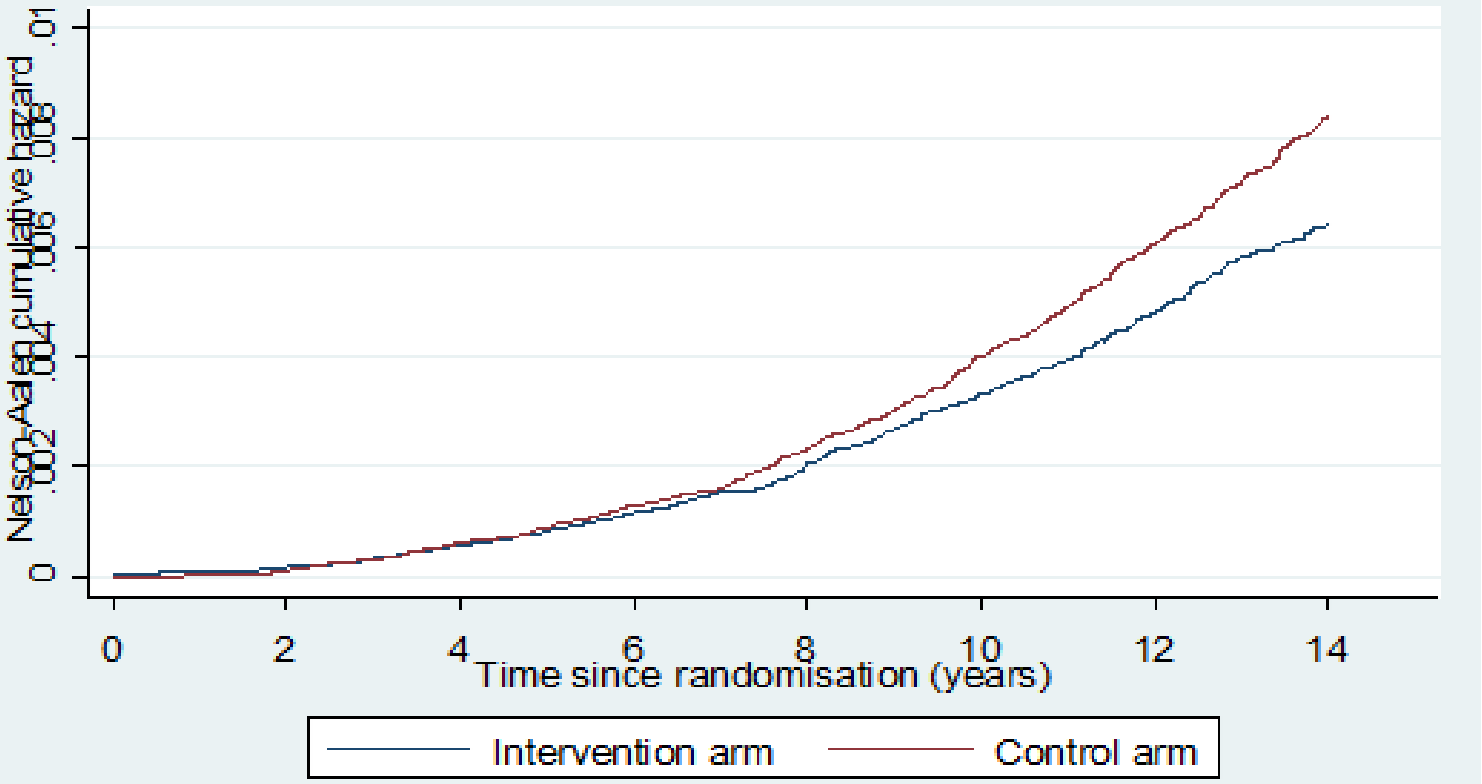
4. LCM results were then used to correct study results for variability in adjudication

European screening trial

Country	Men	Start	Design	Ages	Interval	PC deaths
Finland	80,379	1996	Population	55-69	4	376
Netherlands	34,833	1993	Volunteer	55-74	4	166
Sweden	11,852	1995	Population	50-64	2	109
Italy	14,517	1996	Volunteer	55-70	4	41
Belgium	8562	1991	Volunteer	55-74	4	47
Spain	2197	1996	Volunteer	50-69	4	3
Switzerland	9903	1998	Volunteer	55-69	4	19
France	79,014	2000	Population	55-69	4	53


Randomised
162,243 men
Screening arm
72,891 men
Control arm
89,352 men
Non-participants
N=12,647
Screened
N=60,244
Normal PSA
50,167 men
PSA elevated
10,077 men
PC
N = 819
Interval ca
N = 958
Screen-detected PC
N = 4951
PC
N = 5396

ERSPC prostate cancer mortality



Adjudication

Medical records for deceased men with PC diagnoses were obtained…

- **CT and x-ray images, PSA results, details of co-morbidity…**
- **Records were anonymized.**
- **Method of cancer detection was removed.**

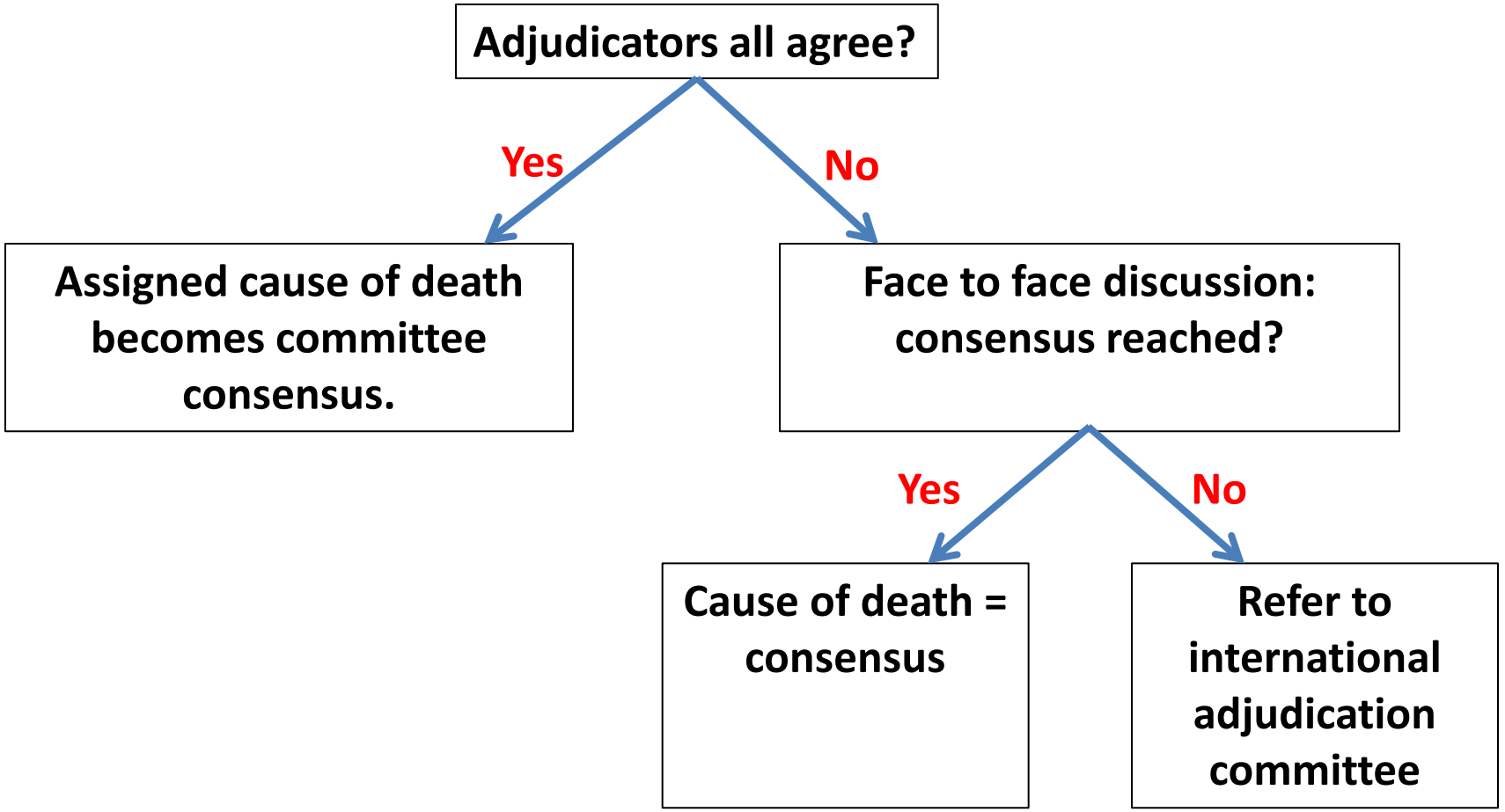
Adjudication

Each country had an adjudication committee:

- **at least three members**
- **not otherwise involved in the trial**
- **representing several medical specialties**
  (commonly urology, pathology and internal medicine).

Each adjudicator assigned a cause of death independently

Adjudication



Analysis

- Descriptive analyses and cross-tabulations
- Agreement between adjudicators
  - Pairwise kappa statistics;
  - McNemar's test for symmetry
- Latent class modelling…

Analysis

Latent class models (LCMs):

LCM's recognise the lack of a gold standard

LCM's formulate the probabilities for a given set of adjudications, conditional on each assumed category for the true (but unknown) cause of death being correct.

These probabilities are then weighted according to the corresponding prevalences (also estimated) of each true cause of death category → MLE's of log-linear model parameters

Parameters of interest:

- **Accuracy (sensitivity and specificity) of adjudicators;**
- **True prevalence of PC death, by study arm (screened vs control);**
- **Association (odds ratios) of the PC death rate *vs.* study arm**

Analysis

Model 1: includes terms {X, T|X, A|X, B|X, C|X…}.

X: true (but unknown) cause of death.

A|X : adjudicator accuracy (reflects conditional probability of adjudicator A recording a particular cause of death, given X). Yields sensitivity and specificity. (Similarly for adjudicators B, C,….)

T|X : prevalence and hence the association between the study arm (T) and PC death rate.

Model 2: same as model 1, but with constraints A|X = B|X = C|X….

Implies that adjudicators have equal sensitivity and equal specificity

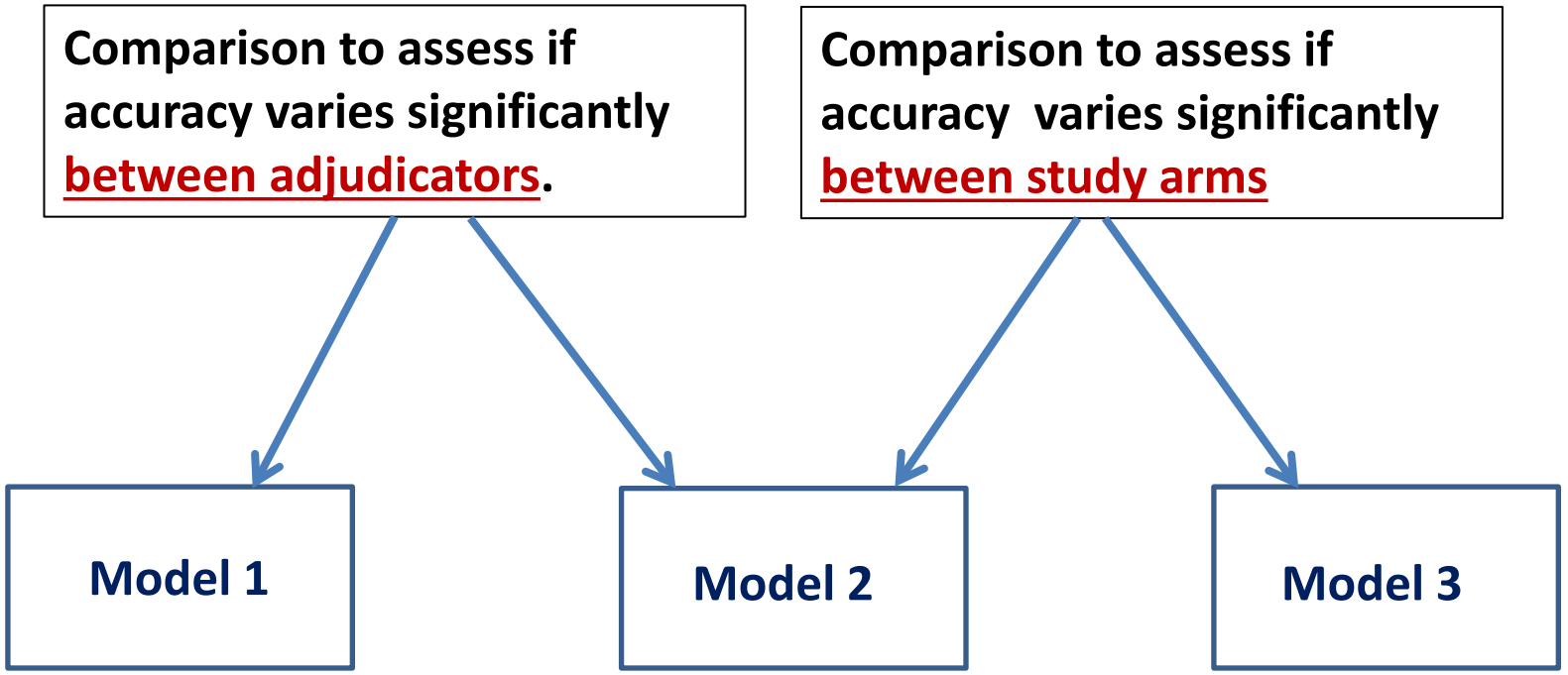
Model 3: includes terms {X, T|X, A|XT, B|XT, C|XT…},

Terms such as A|XT allow accuracy to depend on the study arm.

Analysis

Maximum likelihood estimates of model parameters are found.

Likelihood Ratio statistics used to compare models.



Sample sizes and numbers of deaths, by adjudicator and country

		Adjudicator				
Country	Total cases	1	2	3	4	5
Netherlands	697	659 (95%)	284 (41%)	689 (99%)	16 (2%)	400 (57%)
Belgium	368	368 (100%)	233 (63%)	367 (100%)	367 (100%)	-
Sweden	418	418 (100%)	412 (99%)	415 (99%)	-	-
Finland	435	435 (100%)	435 (100%)	435 (100%)	-	-
Italy	51	51 (100%)	51 (100%)	51 (100%)	51 (100%)	51 (100%)
Switzerland	87	51 (59%)	87 (100%)	87 (100%)	36 (41%)	-

Pairwise agreement (κ statistic) between adjudicators, by country

		Adjudicator pair					
Country	Study arm	1	2	3	4	5	6
Netherlands	Screen	0.92	0.85*	0.94	0.81	0.88*	-
	Control	0.89	0.87	0.93	0.84	0.82	-
	Overall	0.91	0.87*	0.94	0.84	0.86	-
Belgium	Screen	0.92	0.89	0.86	0.93	0.89	0.92
	Control	0.89	0.89	0.92	0.91	0.97	0.92
	Overall	0.91	0.89	0.89	0.92	0.93	0.93
Sweden	Screen	0.95	0.93	0.97	-	-	-
	Control	0.94	0.89	0.90	-	-	-
	Overall	0.94	0.91	0.94	-	-	-
Finland	Screen	0.90	0.85	0.89	-	-	-
	Control	0.89	0.86*	0.92*	-	-	-
	Overall	0.89	0.86*	0.91	-	-	-
Switzerland	Screen	0.78	0.69	0.73	0.92	0.83	
	Control	0.31	0.49*	0.75	1.00	1.00	
	Overall	0.56	0.61	0.74	0.93	0.86	

*: indicates $p < 0.05$ on McNemar 2-sided test for symmetry.

Estimated false positive and false negative adjudication rates (%) by adjudicator, overall, and by study arm

		Data source						
Country	Error rate	Adj. #1	Adj. #2	Adj. #3	Adj. #4	Overall	Screening arm	Control arm
Netherlands	FPR(%)	0.4	0.5	1.8	0.5	0.9	0.7	1.3
	FNR(%)	10.4	0.0	3.5	10.0	7.0	7.4	6.4
Belgium	FPR(%)	0.7	1.1	0.7	0.0	0.6	0.5	0.6
	FNR(%)	4.5	5.9	7.4	6.0	6.0	7.7	4.7
Sweden	FPR(%)	0.8	0.8	2.8	-	1.5	1.7	2.3
	FNR(%)	3.7	0.6	1.3	-	1.9	0.6	3.1
Finland	FPR(%)	1.9	1.2	6.2	-	2.5	2.4	2.7
	FNR(%)	4.9	1.5	1.2	-	3.1	3.3	3.1
Switzerland	FPR(%)	20.8	5.6	4.4	0.0	6.9	2.2	20.7
	FNR(%)	6.2	5.9	10.4	0.0	7.5	10.7	0.0

Results for individual adjudicators are from model 1; the overall results are from model 2; the screening and control arm results are from model 3.

Estimated false positive and false negative adjudication rates (%) by adjudicator, overall, and by study arm

		Data source						
Country	Error rate	Adj. #1	Adj. #2	Adj. #3	Adj. #4	Overall	Screening arm	Control arm
Netherlands	FPR(%)	0.4	0.5	1.8	0.5	0.9	0.7	1.3
	FNR(%)	10.4	0.0	3.5	10.0	7.0	7.4	6.4
Belgium	FPR(%)	0.7	1.1	0.7	0.0	0.6	0.5	0.6
	FNR(%)	4.5	5.9	7.4	6.0	6.0	7.7	4.7
Sweden	FPR(%)	0.8	0.8	2.8	-	1.5	1.7	2.3
	FNR(%)	3.7	0.6	1.3	-	1.9	0.6	3.1
Finland	FPR(%)	1.9	1.2	6.2	-	2.5	2.4	2.7
	FNR(%)	4.9	1.5	1.2	-	3.1	3.3	3.1
Switzerland	FPR(%)	20.8	5.6	4.4	0.0	6.9	2.2	20.7
	FNR(%)	6.2	5.9	10.4	0.0	7.5	10.7	0.0

Results for individual adjudicators are from model 1; the overall results are from model 2; the screening and control arm results are from model 3.

Estimated false positive and false negative adjudication rates (%) by adjudicator, overall, and by study arm

		Data source						
Country	Error rate	Adj. #1	Adj. #2	Adj. #3	Adj. #4	Overall	Screening arm	Control arm
Netherlands	FPR(%)	0.4	0.5	1.8	0.5	**0.9**	0.7	1.3
	FNR(%)	10.4	0.0	3.5	10.0	**7.0**	7.4	6.4
Belgium	FPR(%)	0.7	1.1	0.7	0.0	**0.6**	0.5	0.6
	FNR(%)	4.5	5.9	7.4	6.0	**6.0**	7.7	4.7
Sweden	FPR(%)	0.8	0.8	2.8	-	**1.5**	1.7	2.3
	FNR(%)	3.7	0.6	1.3	-	**1.9**	0.6	3.1
Finland	FPR(%)	1.9	1.2	6.2	-	**2.5**	2.4	2.7
	FNR(%)	4.9	1.5	1.2	-	**3.1**	3.3	3.1
Switzerland	FPR(%)	20.8	5.6	4.4	0.0	**6.9**	2.2	20.7
	FNR(%)	6.2	5.9	10.4	0.0	**7.5**	10.7	0.0

Results for individual adjudicators are from model 1; the overall results are from model 2; the screening and control arm results are from model 3.

Estimated false positive and false negative adjudication rates (%) by adjudicator, overall, and by study arm

		Data source						
Country	Error rate	Adj. #1	Adj. #2	Adj. #3	Adj. #4	Overall	Screening arm	Control arm
Netherlands	FPR(%)	0.4	0.5	1.8	0.5	0.9	**0.7**	**1.3**
	FNR(%)	10.4	0.0	3.5	10.0	7.0	**7.4**	**6.4**
Belgium	FPR(%)	0.7	1.1	0.7	0.0	0.6	**0.5**	**0.6**
	FNR(%)	4.5	5.9	7.4	6.0	6.0	**7.7**	**4.7**
Sweden	FPR(%)	0.8	0.8	2.8	-	1.5	**1.7**	**2.3**
	FNR(%)	3.7	0.6	1.3	-	1.9	**0.6**	**3.1**
Finland	FPR(%)	1.9	1.2	6.2	-	2.5	**2.4**	**2.7**
	FNR(%)	4.9	1.5	1.2	-	3.1	**3.3**	**3.1**
Switzerland	FPR(%)	20.8	5.6	4.4	0.0	6.9	**2.2**	**20.7**
	FNR(%)	6.2	5.9	10.4	0.0	7.5	**10.7**	**0.0**

Results for individual adjudicators are from model 1; the overall results are from model 2; the screening and control arm results are from model 3.

Likelihood ratio tests of heterogeneity in adjudication accuracy

Country	Test of adjudicator heterogeneity (Latent class models 1 vs. 2)			Test of study arm heterogeneity (Latent class models 2 vs. 3)		
	LR statistic	D.f.	p	LR statistic	D.f.	p
Netherlands	20.84	6	< 0.01	0.80	2	0.67
Belgium	4.78	6	0.57	0.90	2	0.64
Sweden	8.54	4	0.07	6.24	2	0.04
Finland	15.62	4	< 0.01	0.04	2	0.98
Switzerland	11.98	6	0.06	10.58	2	<0.01

Likelihood ratio tests of heterogeneity in adjudication accuracy

Country	Test of adjudicator heterogeneity (Latent class models 1 vs. 2)			Test of study arm heterogeneity (Latent class models 2 vs. 3)		
	LR statistic	D.f.	p	LR statistic	D.f.	p
Netherlands	**20.84**	**6**	**< 0.01**	0.80	2	0.67
Belgium	**4.78**	**6**	**0.57**	0.90	2	0.64
Sweden	**8.54**	**4**	**0.07**	6.24	2	0.04
Finland	**15.62**	**4**	**< 0.01**	0.04	2	0.98
Switzerland	**11.98**	**6**	**0.06**	10.58	2	<0.01

Likelihood ratio tests of heterogeneity in adjudication accuracy

Country	Test of adjudicator heterogeneity (Latent class models 1 vs. 2)			Test of study arm heterogeneity (Latent class models 2 vs. 3)		
	LR statistic	D.f.	p	LR statistic	D.f.	p
Netherlands	20.84	6	< 0.01	**0.80**	**2**	**0.67**
Belgium	4.78	6	0.57	**0.90**	**2**	**0.64**
Sweden	8.54	4	0.07	**6.24**	**2**	**0.04**
Finland	15.62	4	< 0.01	**0.04**	**2**	**0.98**
Switzerland	11.98	6	0.06	**10.58**	**2**	**<0.01**

Estimated odds ratios between prostate cancer death and study arm

Country	Estimation method			
	Empirical[a]	Empirical, corrected using overall estimates of adjudicator accuracy[b]	Empirical, corrected using differential estimates of adjudicator accuracy by study arm[c]	Directly from latent class model[d]
Netherlands	0.342	0.320	0.333	0.332
Belgium	0.759	0.752	0.788	0.902
Sweden	0.355	0.341	0.328	0.368
Finland	0.520	0.498	0.504	0.531
Switzerland	0.625	0.569	1.311	0.437

a: Mantel-Haenszel estimate, consensus by study arm.
b: From LCM 2.
c: From LCM 3.
d: From LCM 2.

Results summary

- Some variation between adjudicators (expected in practice), but pairwise agreement was generally good.
- Only limited evidence of asymmetry between adjudicators.
- No consistent evidence of differential accuracy by trial arm. We conclude that systematic bias arising for this reason was unlikely.
- Model-based and empirical estimates of study OR's were quite similar.

ERSPC vs. PLCO

	ERSPC	PLCO
RR	***0.80 (0.7-0.9)***	***1.12 (1.1-1.2)***
Men	***162,388***	***76,685***
Prior screen	***17%***	***50%***
Contamination	***28%/4yrs***	***40-50%/yr***
FU	***11***	***13***
PC deaths	***761***	***303***
Screened	***82%***	***85%***
Biopsy compliance	***86%***	***32%***

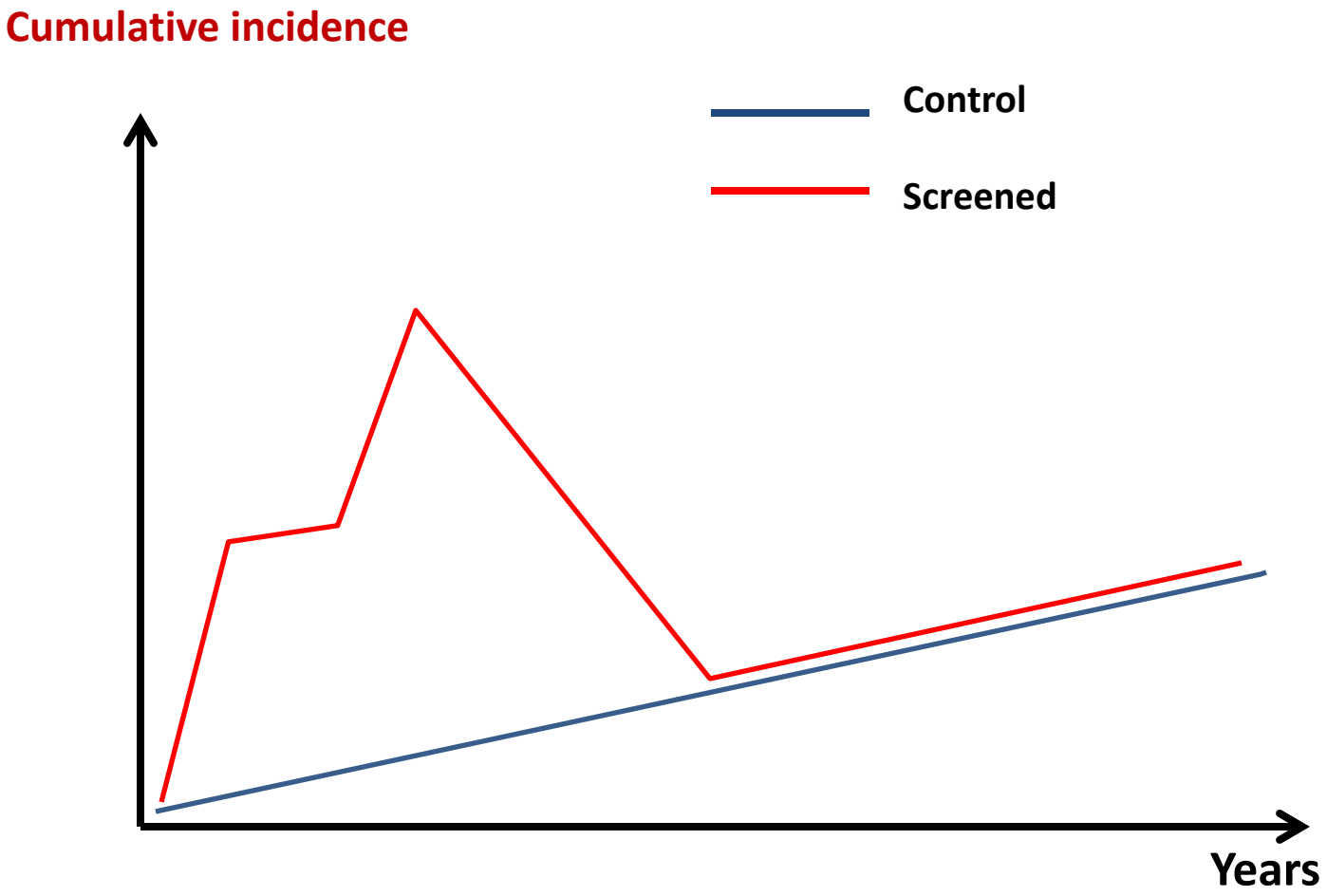
Estimation of the over-diagnosis rate: the catch-up method

The screened group will typically show an excess of cases during the screening period (because of earlier detection).

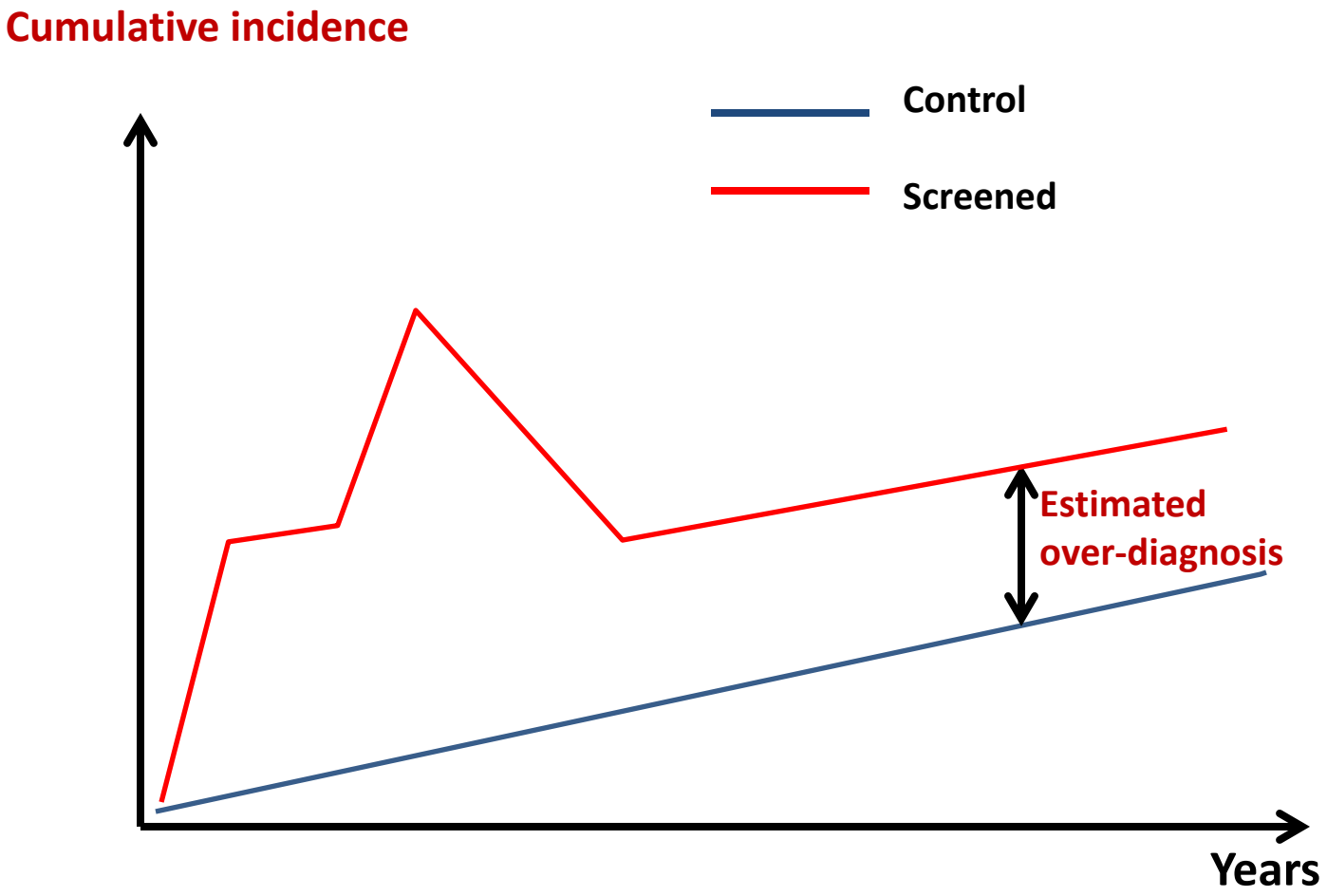
Problem: identify when the number of cases in the control group has "caught up" to the screened group, or where the difference in the cumulative number of cases has stabilised.

Point of stability will indicate the number of overdiagnosed cases.

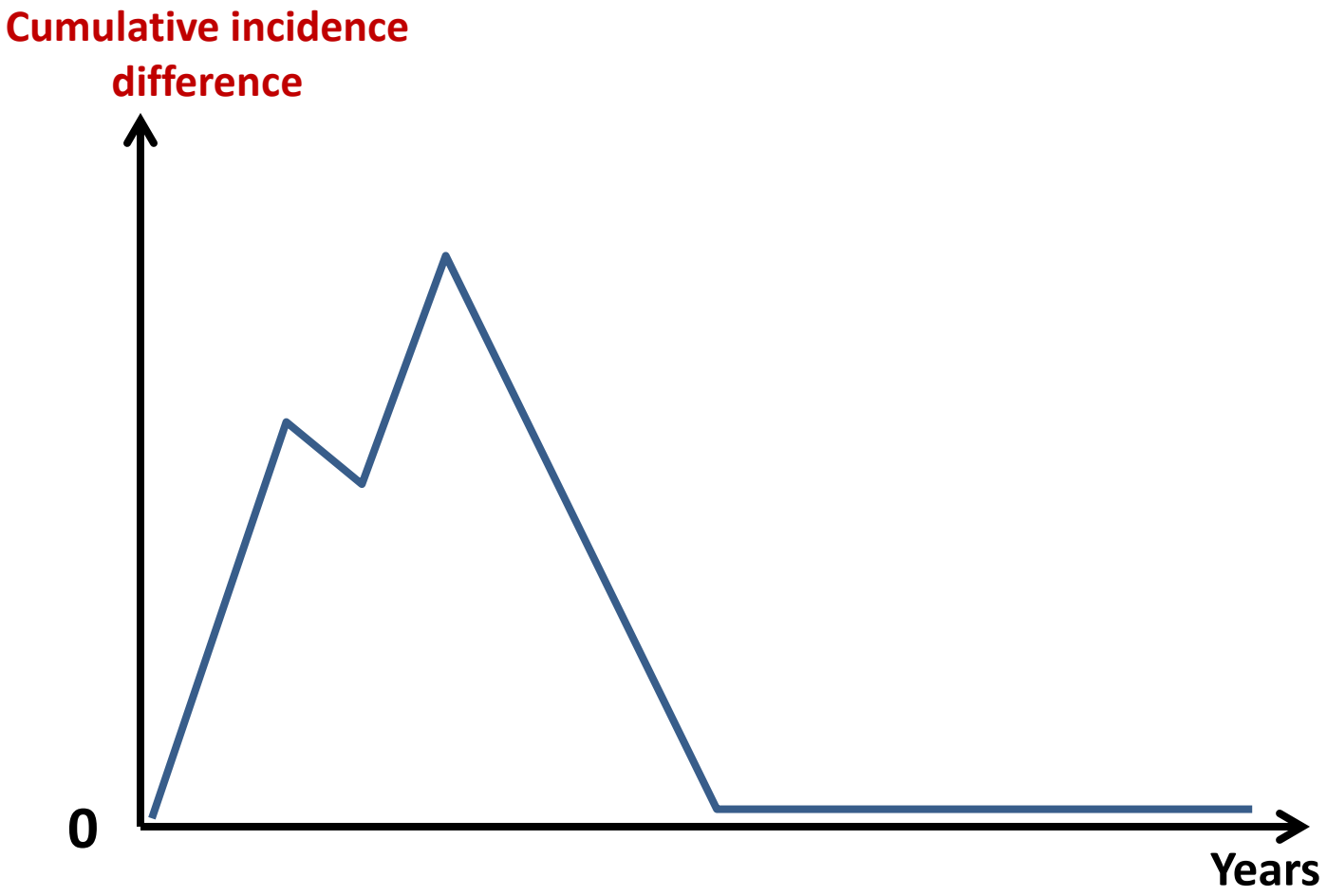
Catch-up method with no over-diagnosis



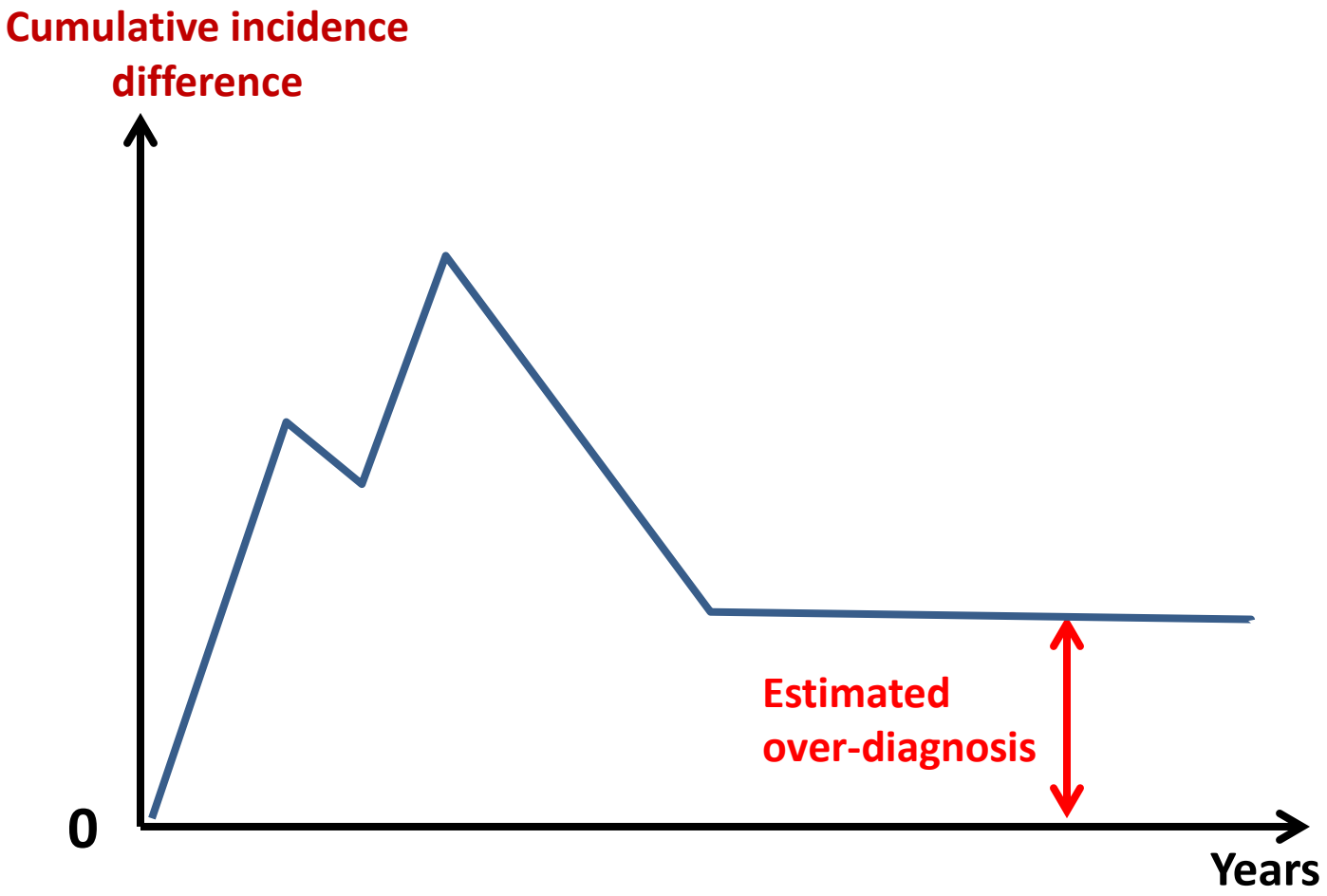
Catch-up method with over-diagnosis



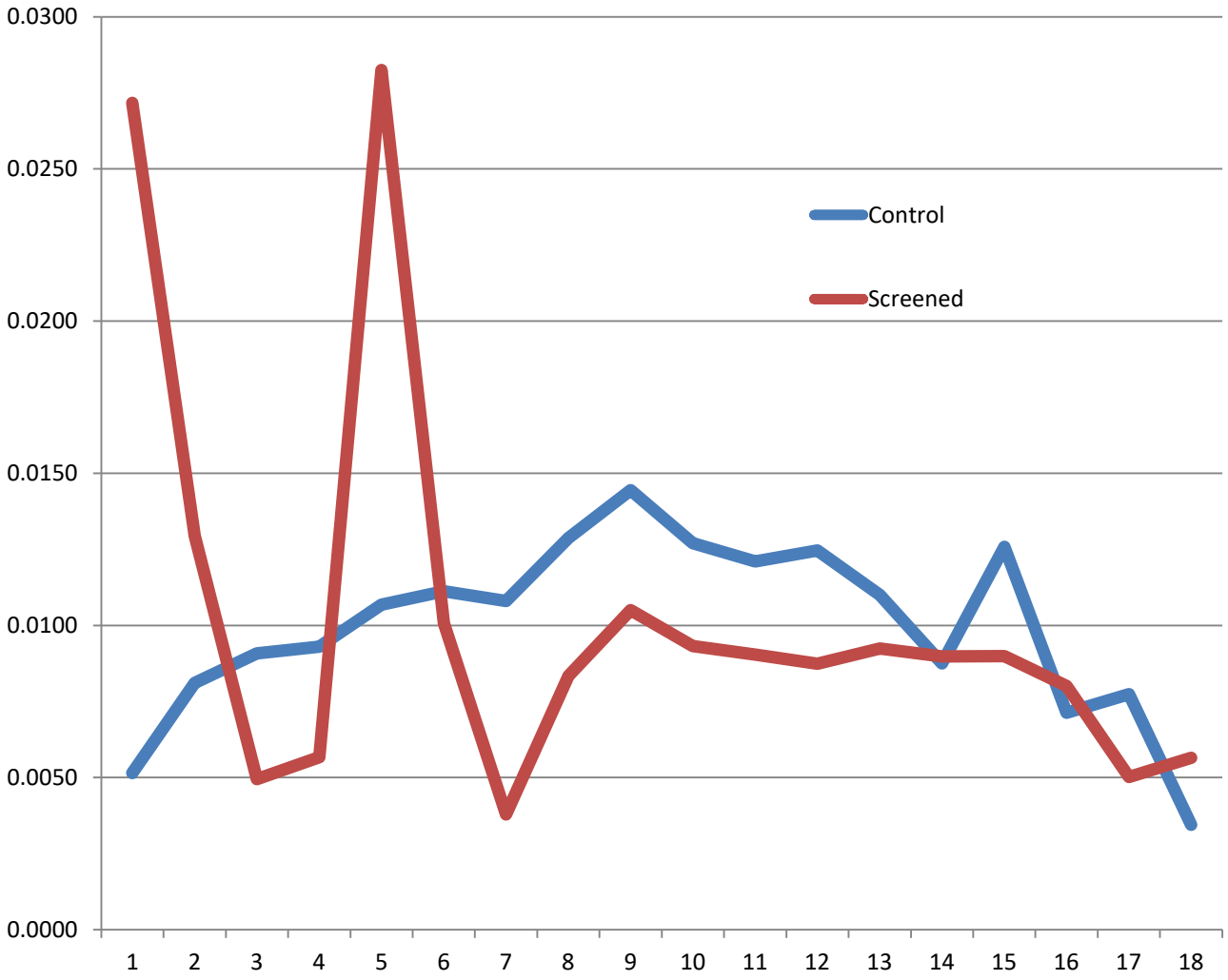
Catch-up method with no over-diagnosis



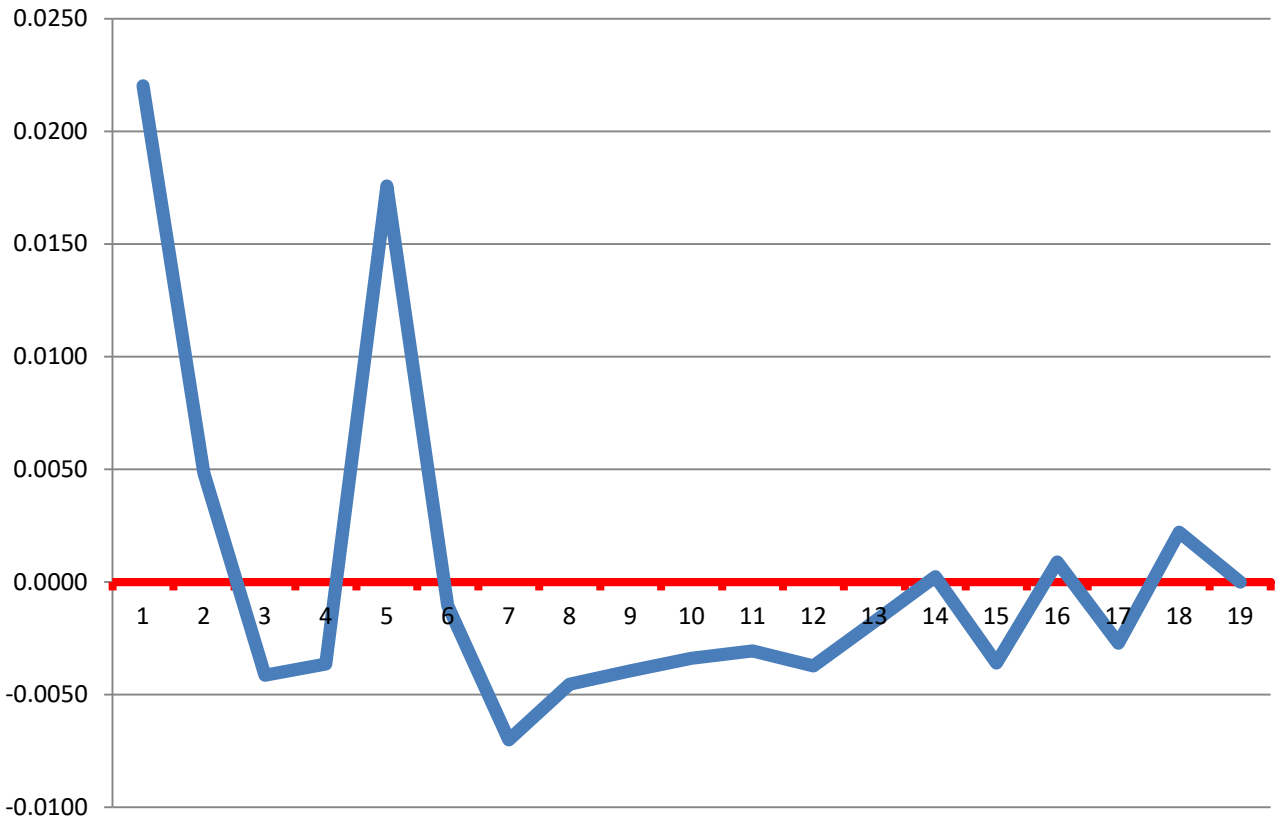
Catch-up method with over-diagnosis



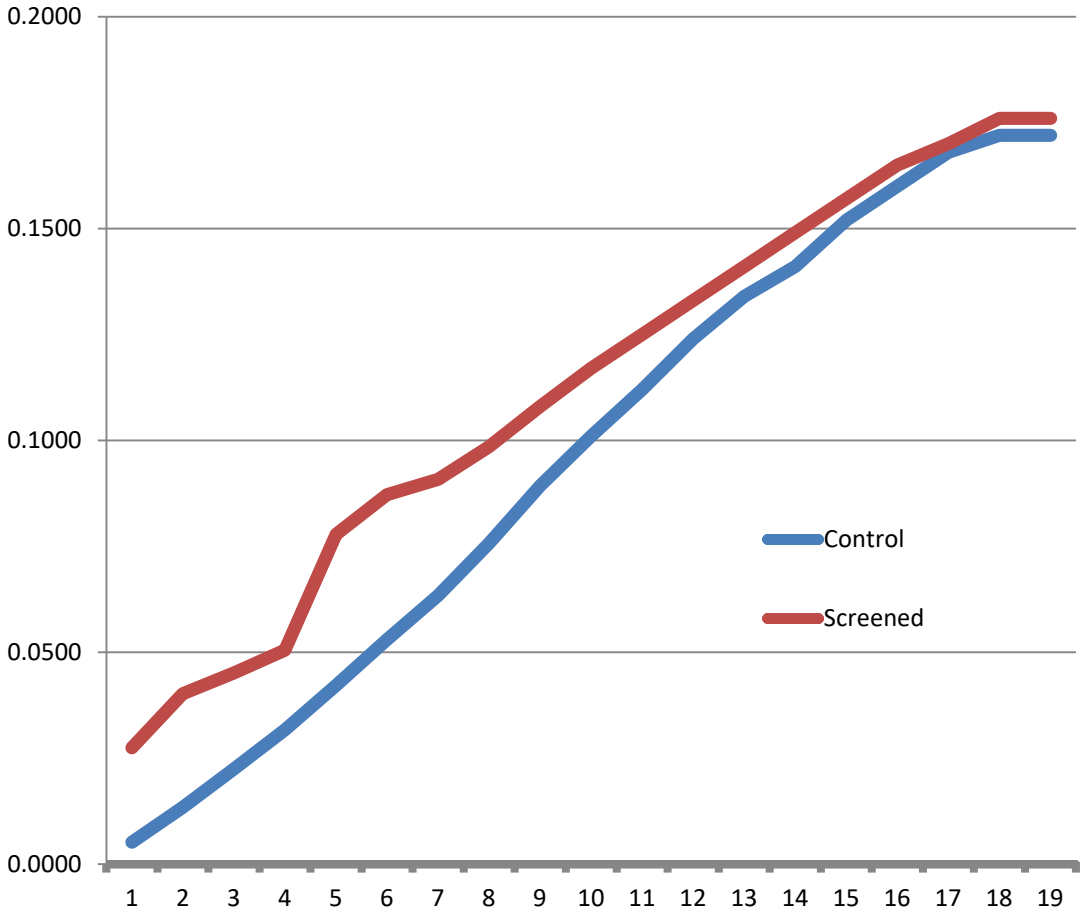
Incidence rates by year (1929-32 cohort)



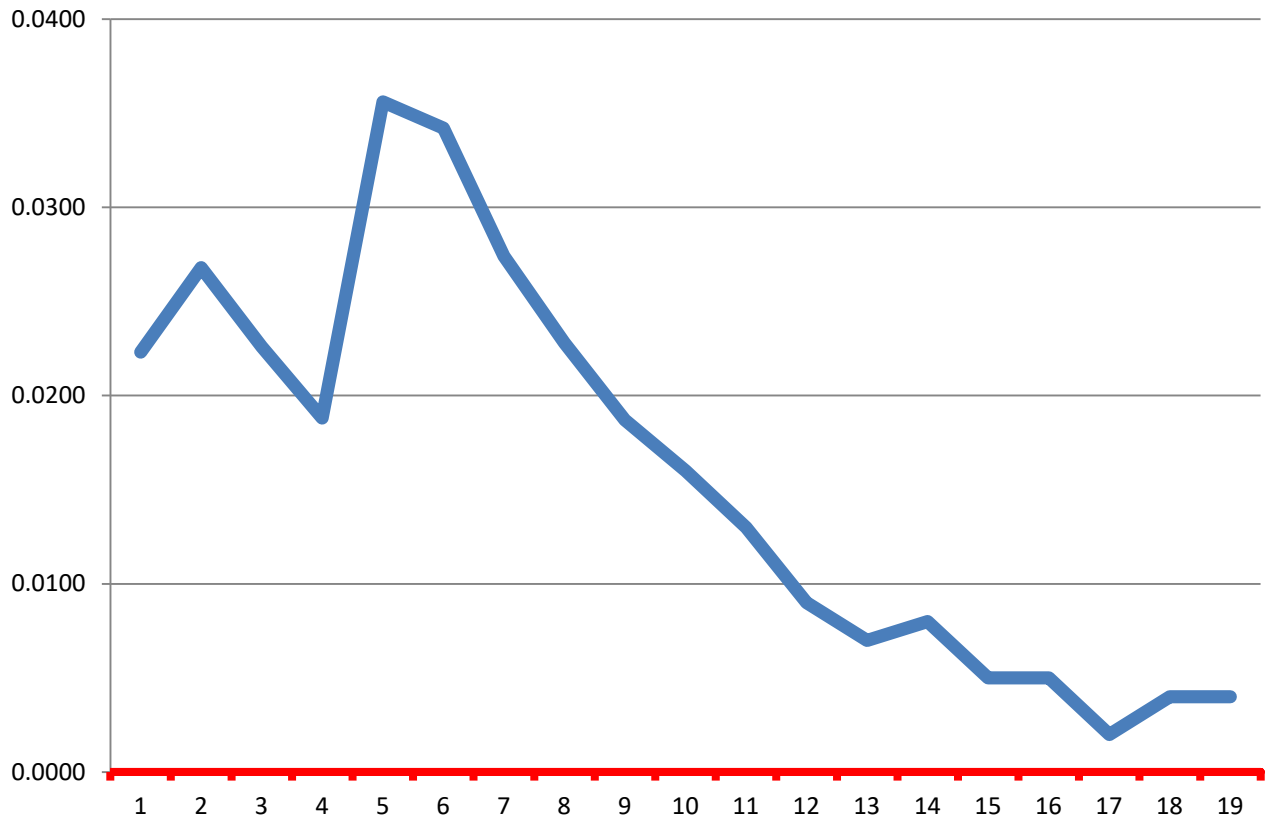
Rate difference by year (1929-32 cohort)



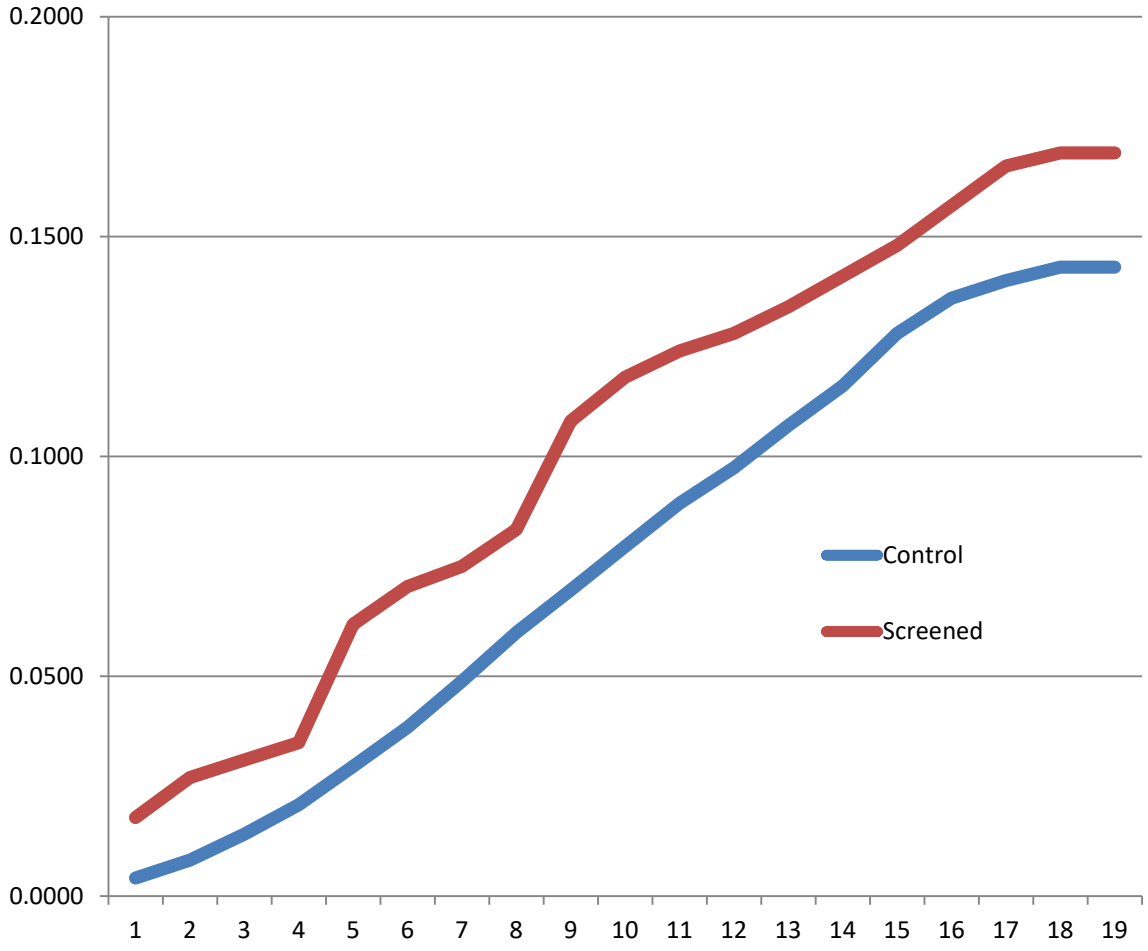
Cumulative incidence rate (1929-32 cohort)



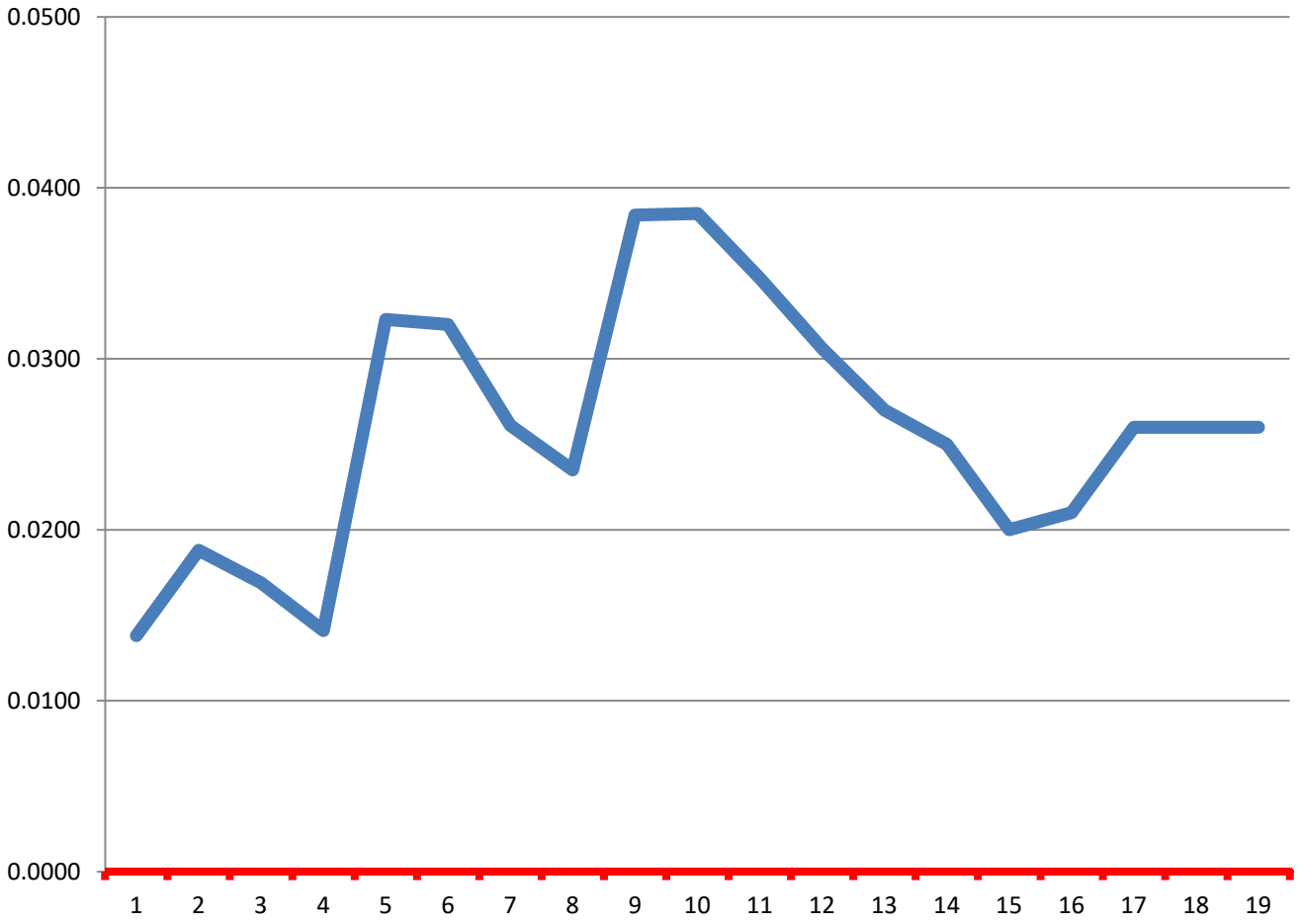
Cumulative incidence difference (1929-32 cohort)



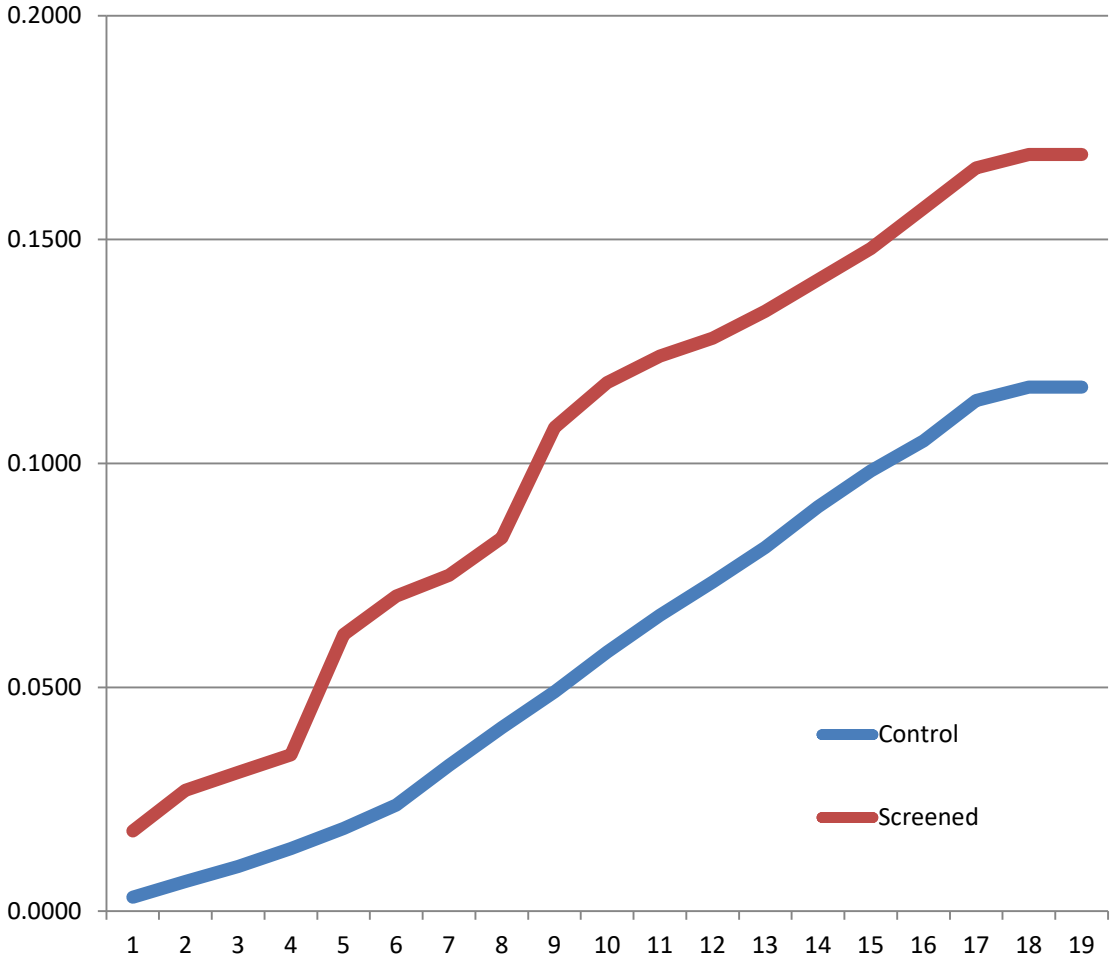
Cumulative incidence rate (1933-36 cohort)



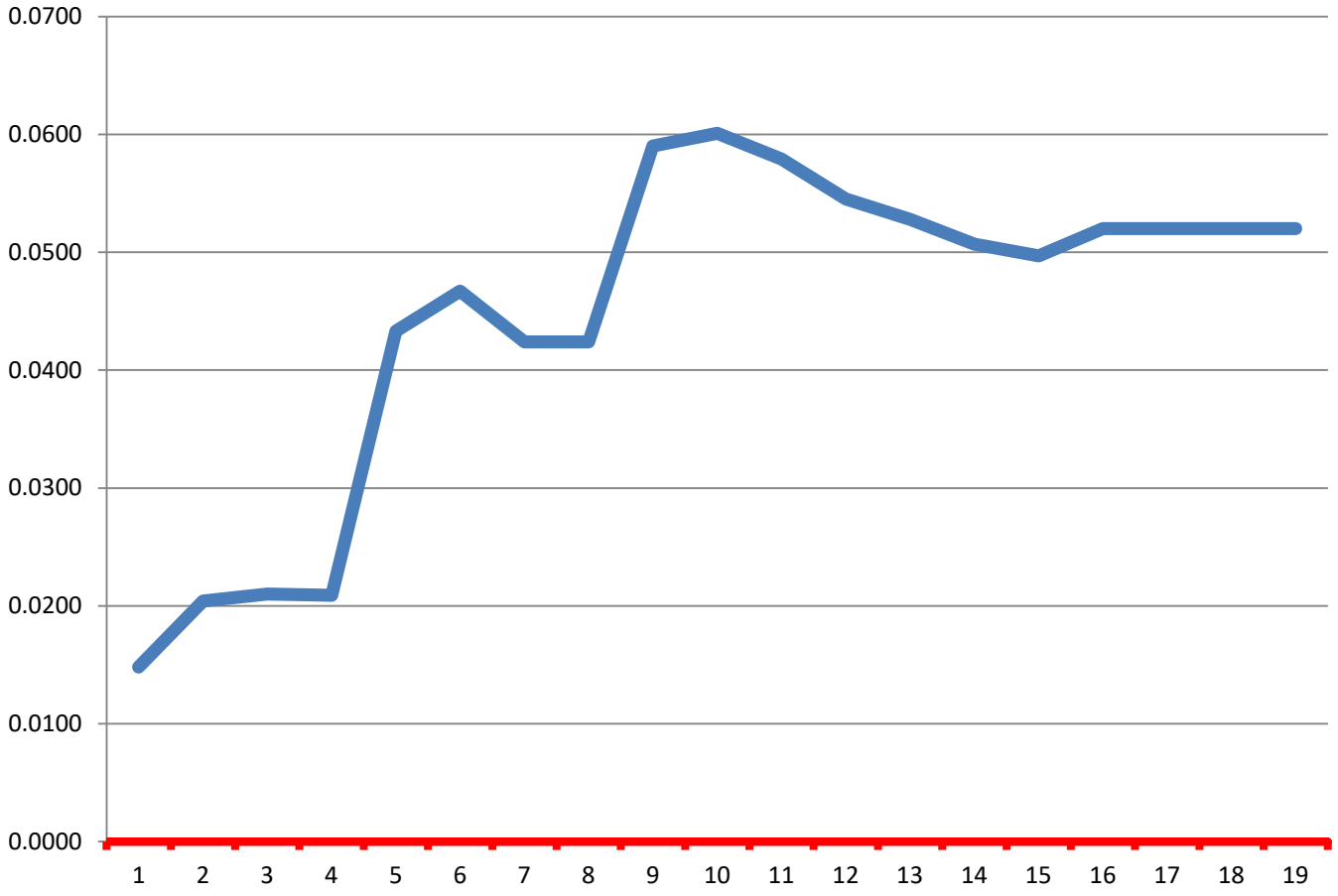
Cumulative incidence difference (1933-36 cohort)



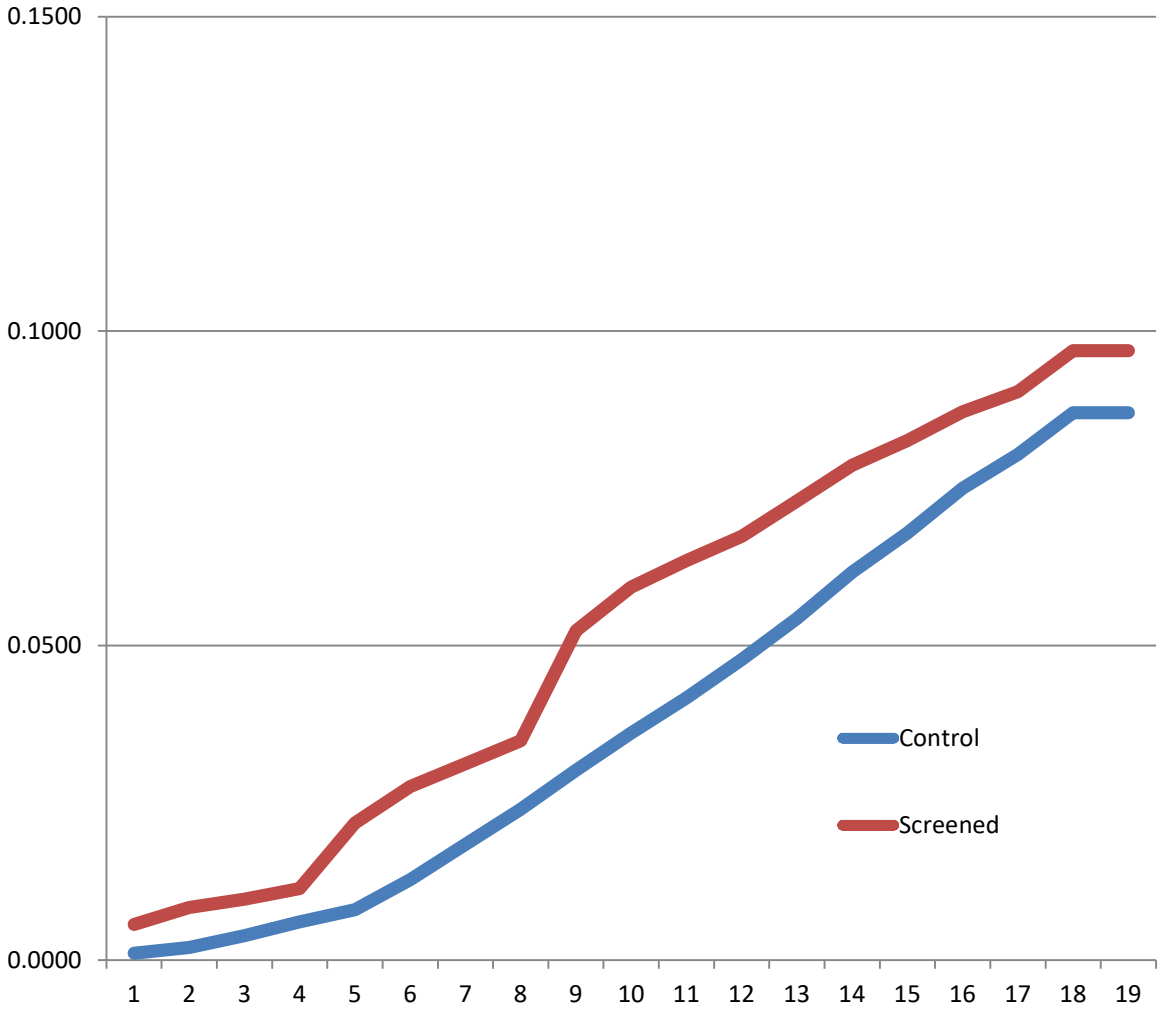
Cumulative incidence rate (1937-40 cohort)



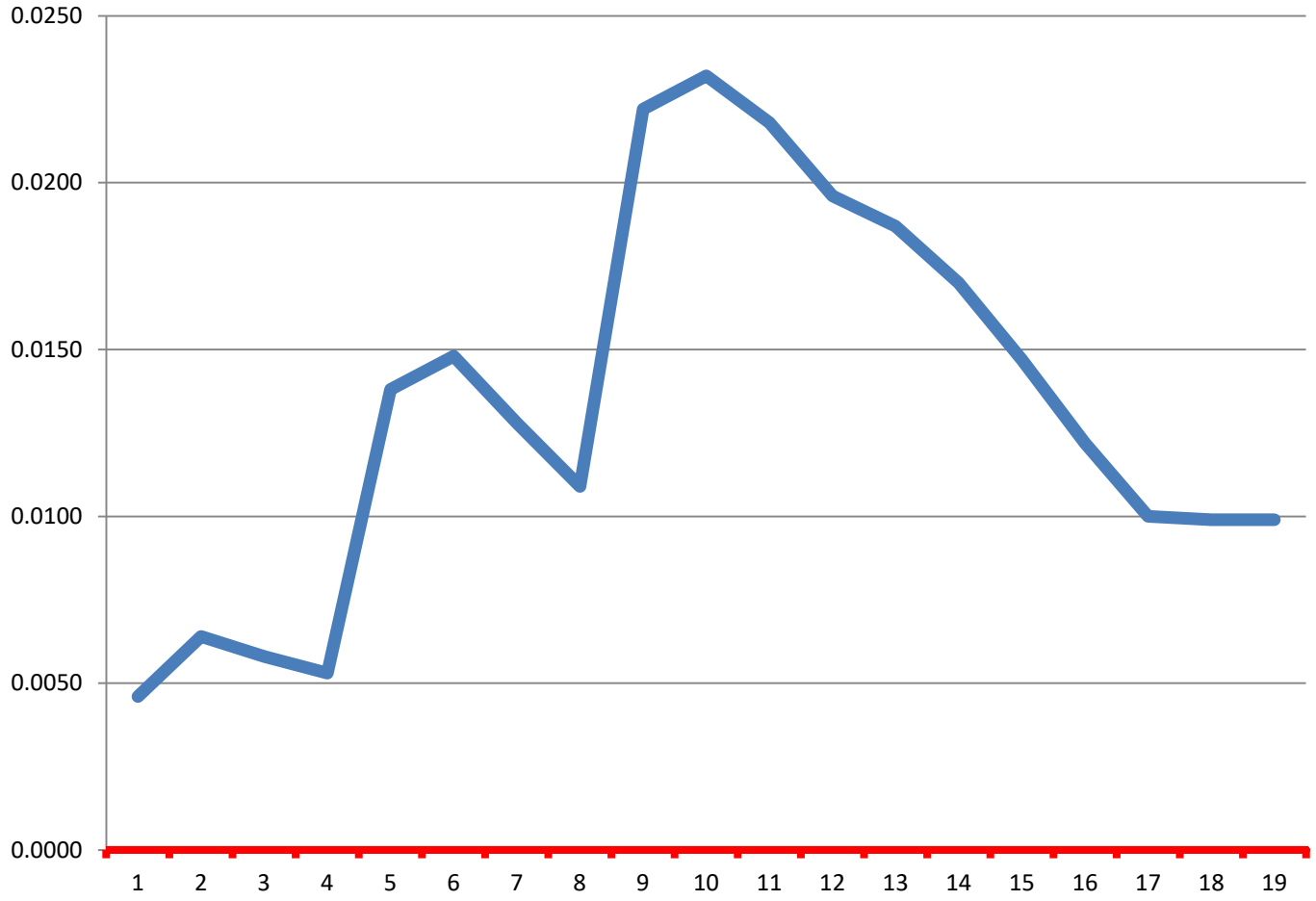
Cumulative incidence difference (1937-40 cohort)



Cumulative incidence rate (1941-44 cohort)

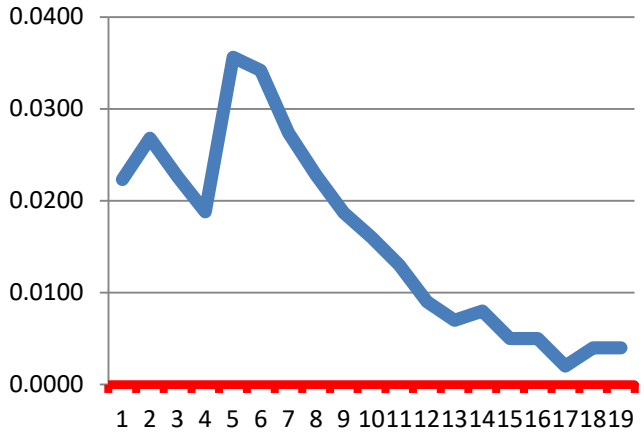


Cumulative incidence difference (1941-44 cohort)

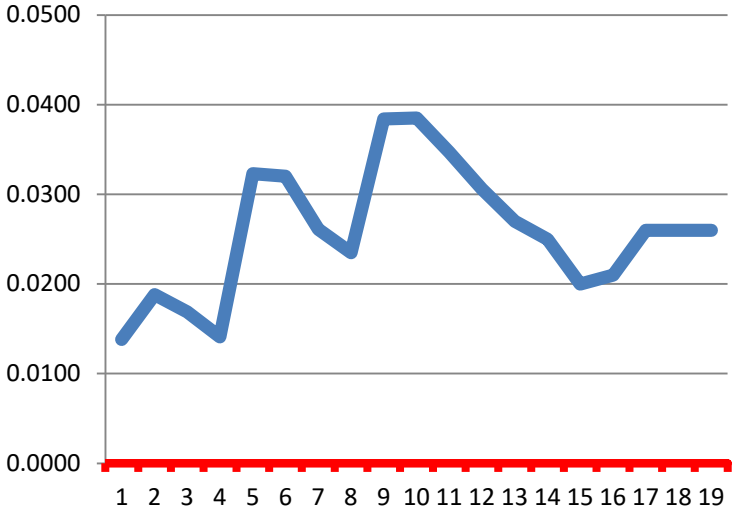


Cumulative incidence differences

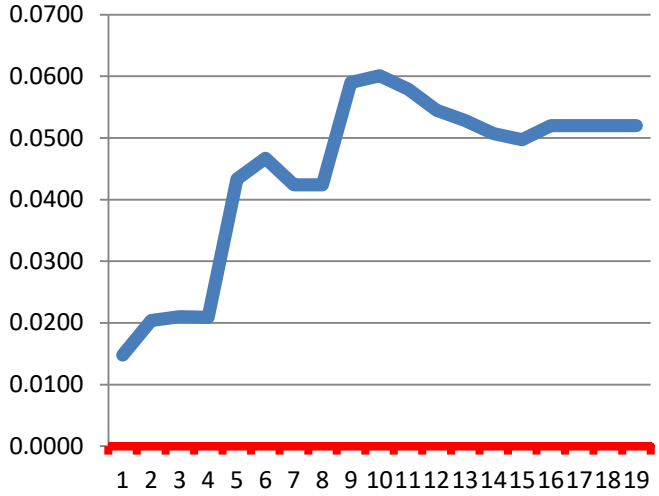
1929-32



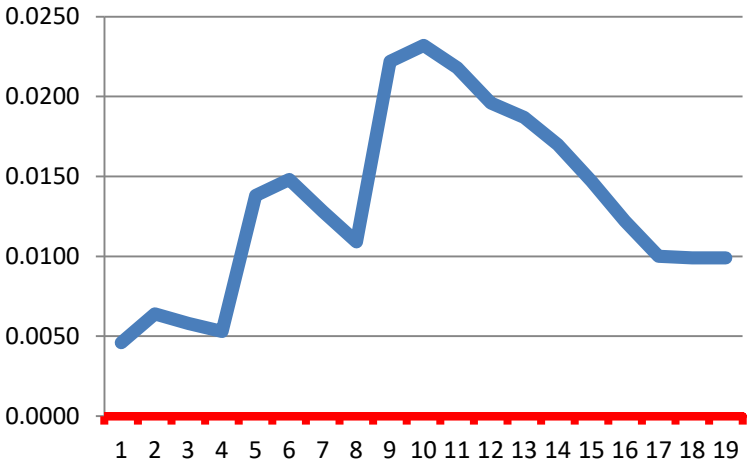
1933-36



1937-40



1941-44



Possible criteria for determining the catch-up point

- **Use year-specific rate differences?**
  - Statistically independent
  - Relatively unstable
  - Heterogeneous (increasing) variance
- **Use cumulative incidence differences?**
  - Increasing statistical dependence
  - Smoother
- **Local measures of variation (expect smaller range, SD)**
  - Width of local window?
- **Local slope (expect zero slope)**
- **LR statistic for suitable piecewise regression model**

Acknowledgements

Hugosson J, de Koning HJ, Talala K, Roobol MJ, Carlsson S, Zappa M, Nelen V, Kwiatkowski M, Paez A, Moss SM, Tammela T, Bangma CH, Aus G, Puliti D, Denis L, Recker F, Randazzo M, Lujan M, Schroder FH, Auvinen A