

Diabetes — an epidemic?
**Some statistical and epidemiological issues in using
pharmaco-epidemiological databases**

Henrik Støvring

hstovring@health.sdu.dk

August 17, 2004

1 Outline of talk

- Scientific issue of primary interest
- Common biases in pharmacoepidemiologic investigations
- A brief review of results regarding the diabetes epidemic

2 Medical Motivation

- Many have reported an increase in prevalence of diabetes and its causes:
 - WHO^a:

“Diabetes is part of the growing epidemic of noncommunicable diseases”

“The good news [...] is that much of the projected increase in diabetes is preventable, through attention to diet and physical activity in the population.”

^a<http://www.who.int/mediacentre/releases/2003/pr86/en/>

– International Diabetes Federation^a:

“The alarming increase of diabetes prevalence is projected to occur because of:

- * Population ageing
- * Unhealthy diet
- * Obesity
- * A sedentary lifestyle”

– Danish Diabetes Association^b:

The increased prevalence of diabetes can *primarily* be explained by changes in lifestyle characterised by reduced physical activity and consumption of an excessively large and fatty diet. [my emphasis]

^a<http://www.idf.org/home/index.cfm?node=264>

^b<http://www.diabetes.dk/wm4450>, my translation

- Conjectured that it fulfills criteria for an epidemic
 - Google search [define:epidemic](#) yields:
 - * widespread outbreak of an infectious disease
 - * disease affecting, or tending to affect, a disproportionately large number of individuals
 - * occurrence of more cases of disease than expected
 - * change in the amount of disease
 - * etc...
- The questions were thus:
 - Did prevalence of diabetes rise?
 - Did incidence of diabetes rise?
- In other words: Give an epidemiologic description of the diabetes population

3 Data

3.1 Geographical region



3.2 Type of information captured

- Data from OPED (1992-1999):
 - Date of redemption
 - ATC-code of dispensed drug
 - * A10: All antidiabetic drugs
 - * A10A: Insulin
 - * A10B: Oral Antidiabetics
 - Civil Registration Number of all subjects, includes date of birth and gender
 - Dates of subjects moving in and out of the county for all subjects
 - Date of death for all subjects

where "all subjects" really means "all subjects present at some point in the County of Fyn during 1992-1999"

4 Statistical issues: sources of bias

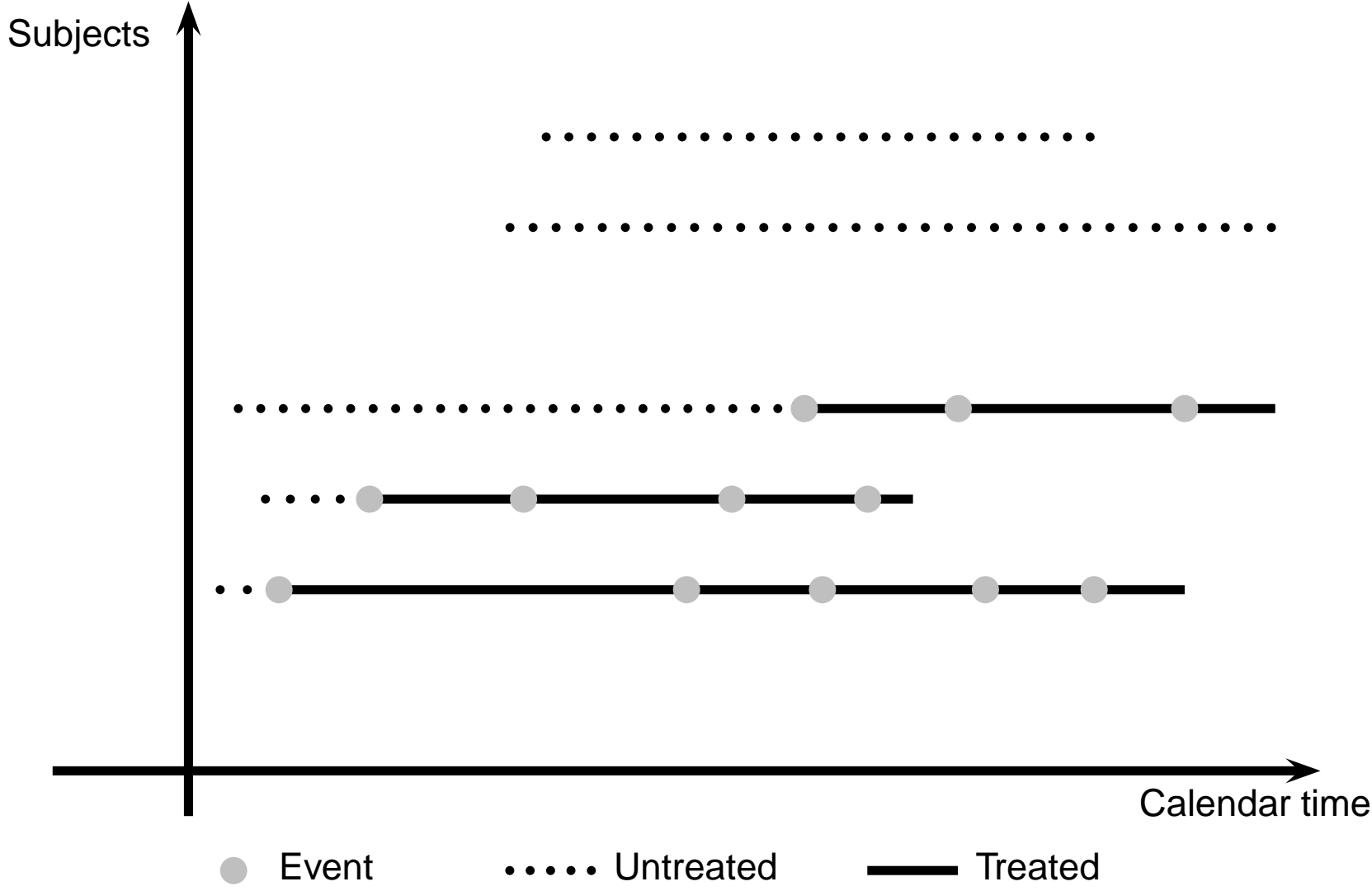
- Misclassification of treatment status
- Exclusion due to migration

5 Standard Method

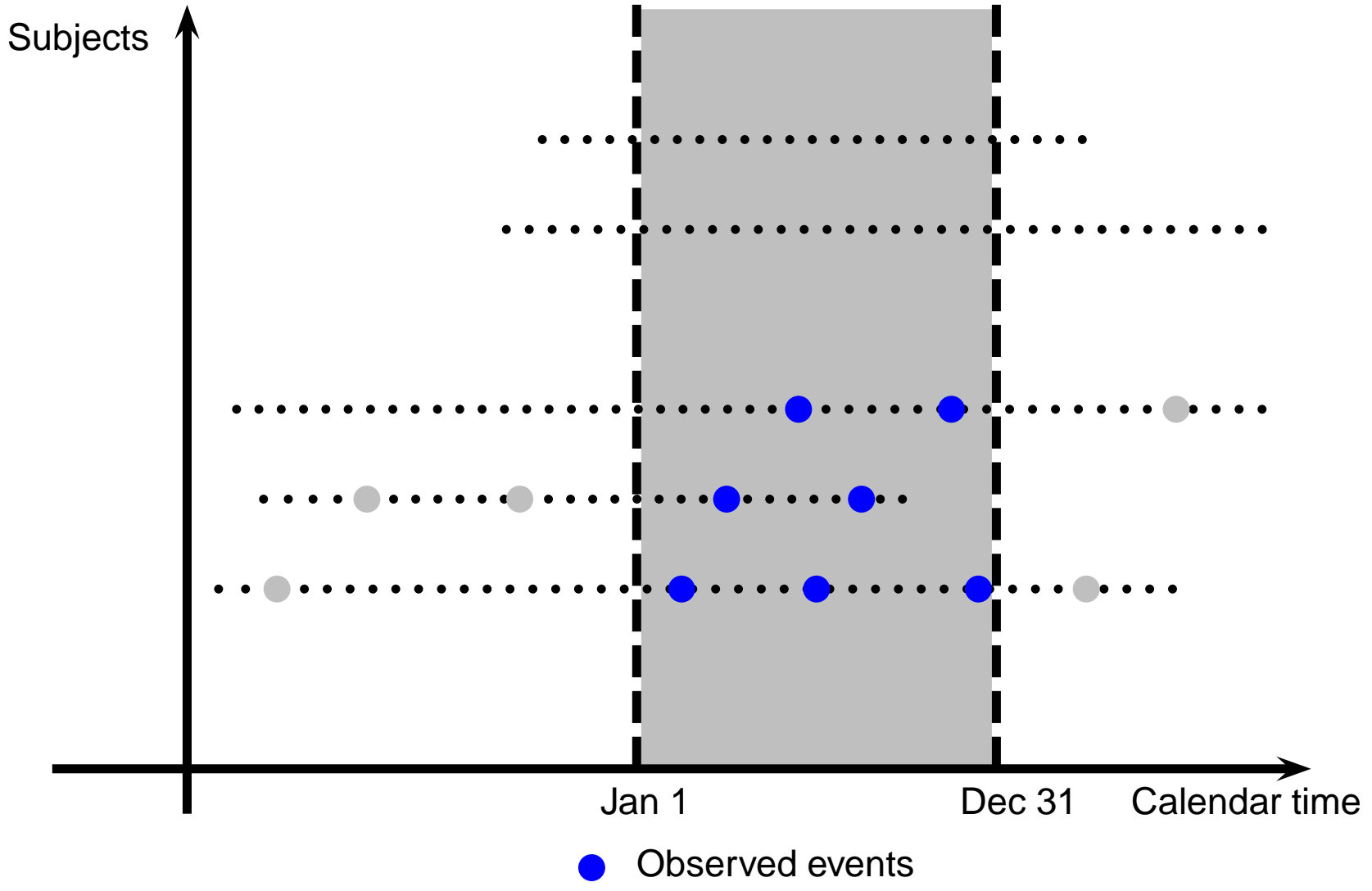
- Identified all subjects present on Jan 1 of each calendar year
- Determined treatment status based on redemptions of antidiabetic drugs in a *run-in period*:
 - No redemptions → untreated
 - At least one redemption → treated
- Note: All immigrating during run-in period were discarded
- One may then compute
 - Prevalence on Jan 1
 - Incidence in the calendar year
 - Mortality in the calendar year among those treated on Jan 1

Examples of use of a run-in period: Støvring et al. 2003; Rahimtoola et al. 2003; Rahimtoola et al. 2002; Mantel-Teeuwisse et al. 2001

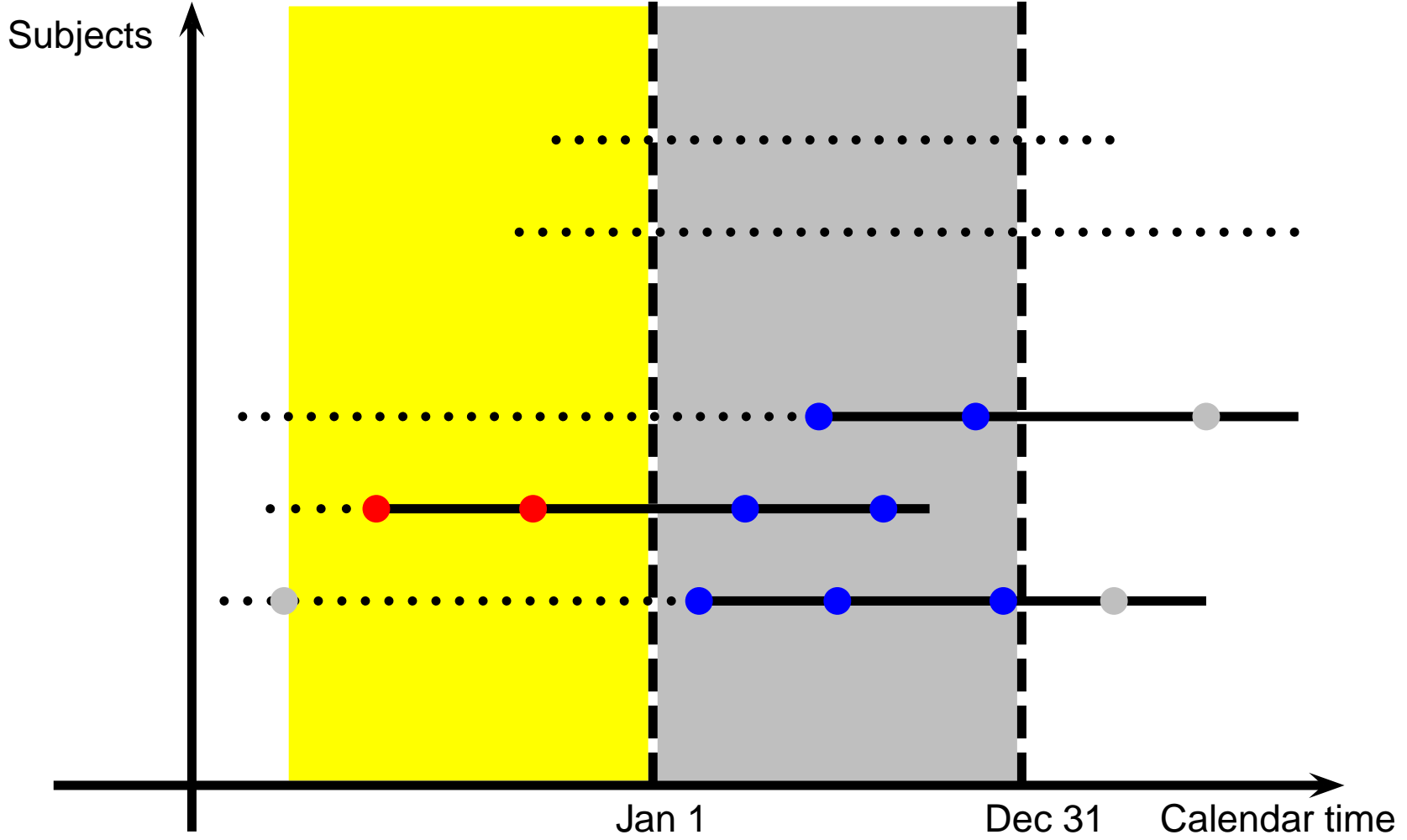
5.1 Run-in period illustrated



5.1.1 What we observe



5.1.2 Using a run-in period



Define treatment status at Jan 1 based on run-in period (yellow area)

5.2 Fundamental issue: choosing the run-in period

- Two strategies:
 1. Dynamic: Maximum available
 2. Static: Fixed length period
 - Clinical judgement
 - Data driven

5.3 Dynamic run-in

- Define a true prevalent at time t as one who
 - has ever had a redemption prior to t
 - is alive at time t (in region of interest)
- The proportion of misclassification depends on length of run-in
- Example:
 - The prevalence on Jan 1, 1999
 - Extend the run-in period successively in one-year increments
 - Converges to true prevalence
 - Result is an increase in prevalence with longer run-in periods

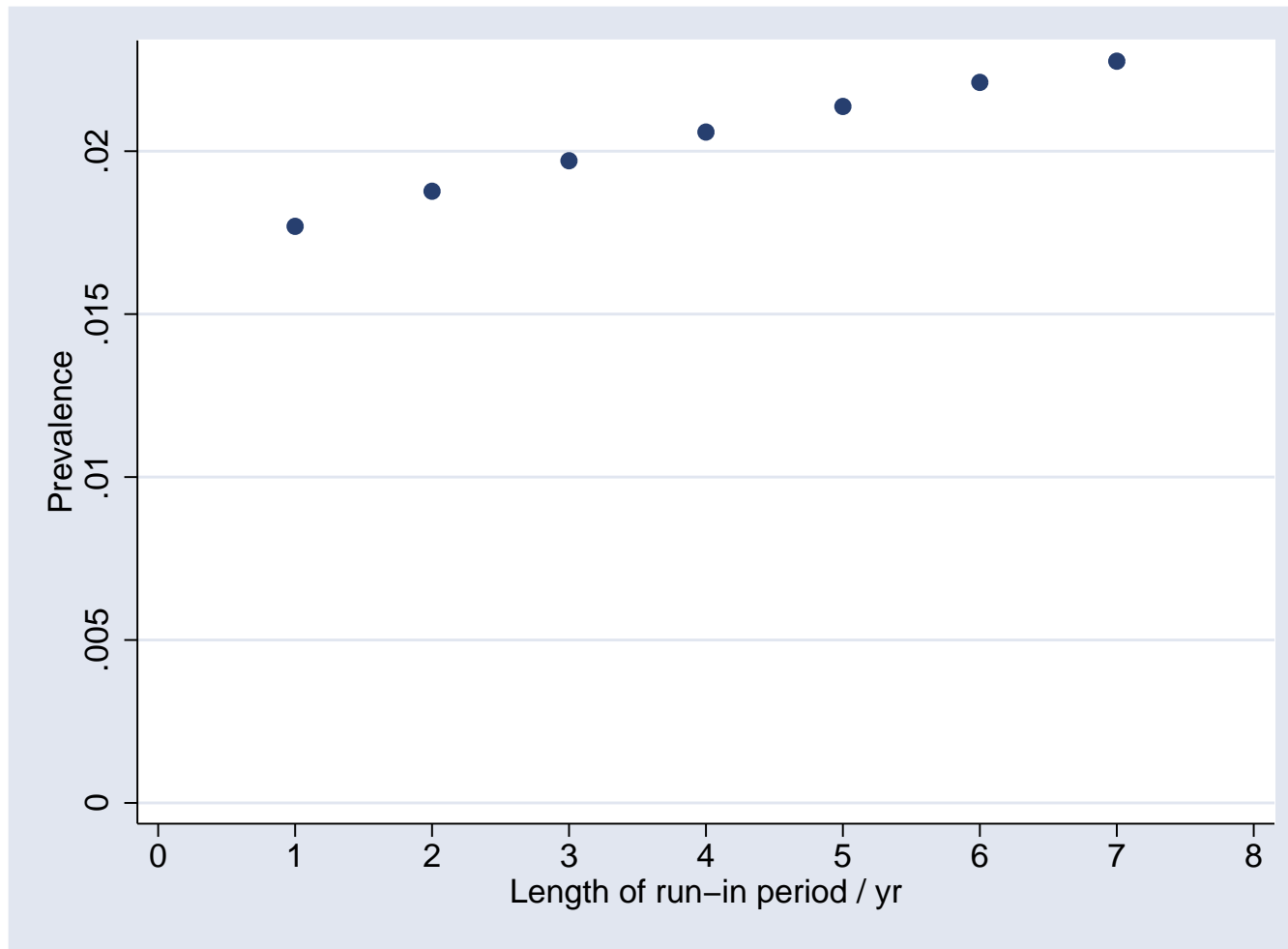


Figure 1: Prevalence of (ever) use of antidiabetics on Jan 1, 1999.

- Using the prevalence with seven years run-in as reference, the relative bias in prevalence with one year run-in is 22.3%
- Complications:
 - Longer run-in period excludes more immigrants (n drops from 453,882 to 374,095)
 - Thus the estimates are not based on the same study population
- Restricting the sample to all not immigrating in the 7 years prior to Jan 1 1999, the relative bias with one year run-in drops to 11.2%

5.3.1 Conclusion regarding dynamic run-in

- Avoids choosing the length of the run-in period
- Secular trends in prevalence will be upward biased
- Secular trends in incidence will be downward biased
- Dynamic run-in periods should not be used when estimating time trends

5.4 Fixed length run-in

- Using clinical judgement, a one year run-in period for antidiabetics is reasonable
- Note that implicitly this corresponds to a change in definition of prevalence:

A subject is prevalent at t , if (s)he is a current user at time t

- With pharmacoepidemiologic databases this is the type of prevalence one can most realistically hope to estimate
- If the proportion of misclassification is constant over time, then estimates of calendar time trends in prevalence are virtually unbiased^a

^aSimulation studies indicate a slight underestimate ($< 0.5\%$ bias) for realistic scenarios

- But: proportion of misclassification may not be constant over time
- During year 2000 new rules for subsidizing drugs was implemented in Denmark
- Did stock piling take place prior to start of the new rules?

5.4.1 Investigating non-constant misclassification

- Consider all those prevalent on Jan 1 of each year
- In other words: all with a prescription on antidiabetics in the previous year
- How many of these had a gap to their next prescription of more than one year?
- That is: how many would we consider to have stopped treatment according to the prevalence definition of being a current user?

Year	One year gap		Two years gap	
	<i>n</i>	%	<i>n</i>	%
1993	198	6.26		
1994	209	6.41	269	7.99
1995	199	5.95	272	7.85
1996	172	5.10	238	6.80
1997	188	5.25	235	6.42
1998	163	4.47	240	6.38
1999	135	3.61	210	5.47
2000	150	3.82	207	5.18
2001	282	6.83	237	5.63
2002	176	4.06	232	5.22
2003	189	4.10	232	4.95

Table 1: Number of women discontinuing antidiabetic treatment

- Marked increase in 2001
- Probably reflects stock piling in 2000
- Two years run-in nearly eliminates the problem

Year	One year gap		Two years gap	
	<i>n</i>	%	<i>n</i>	%
1993	97	7.27		
1994	82	5.92	115	8.02
1995	88	6.16	113	7.72
1996	73	5.05	116	7.73
1997	80	5.27	113	7.28
1998	45	2.90	99	6.16
1999	44	2.81	65	4.09
2000	41	2.48	60	3.60
2001	159	9.05	76	4.31
2002	77	4.17	92	4.90
2003	89	4.50	105	5.24

Table 2: Number of women discontinuing insulin treatment

- Insulin users are more regular users - planning ahead is possible
- Very marked stock piling
- Two years run-in reduces the problem

5.5 Consequences of using a fixed run-in period

- The estimated prevalence is not the classical epidemiologic prevalence of ever use, rather it is an estimate of current use
- Incidence is **not** the classical incidence (first time use ever), rather it is an estimate of transitions from untreated to treated
- In principle, longer run-in periods decreases difference between ever-use prevalence and “current”-use prevalence
- If misclassification is constant over time we get nearly unbiased estimates of secular trends in prevalence
- If misclassification is not constant over time, secular trend estimates may get biased

6 Migration

- Using a run-in period induces selection bias:
- People immigrating to capture area during the run-in period are discarded in the ordinary run-in approach
- If prevalence among migrants differs from prevalence of non-migrants prevalence estimates become biased
- If prevalence among migrants relative to non-migrants are constant, migration proportions are constant and realistic ($\approx 5\%$ per year), then trend estimates for prevalence are virtually unbiased ($\ll 1\%$)

7 Methods not relying on a run-in period

- Idea: look at the distribution from index date to first subsequent prescription (first suggested in (Hallas, Gaist, and Bjerrum 1997))
- Bias problems:
 - Stationarity/independence of prescription renewals and index date
 - Parametric specification
- Computationally intensive means that it is harder to incorporate covariates
- For further details, see (Støvring 2002) and (Støvring and Vach 2004) (available upon request)

8 Epidemiologic analysis – Was there an epidemic of Diabetes in Denmark during the 1990's?

- Used a one year run-in period
- Logistic regression for prevalence
- Poisson regression for incidence and mortality
- Covariates
 - Calendar year
 - Age in eight categories (cut points at 20, 30, . . . , 90)
 - Gender
 - Interaction of Age and Gender
- Robust variance estimates for prevalence and incidence
- Details are in (Støvring, Andersen, Beck-Nielsen, Green, and Vach 2003)

9 Results

9.1 Trend estimates

Treatment	Gender	Prevalence		Incidence	
All Antidiabetic Medications	F	1.026	(1.020; 1.032)	1.010	(.993; 1.027)
	M	1.041	(1.036; 1.047)	1.015	(1.000; 1.031)
Insulin	F	1.024	(1.016; 1.033)	.980	(.954; 1.006)
	M	1.046	(1.038; 1.054)	.997	(.973; 1.022)
Oral Antidiabetic Medications	F	1.028	(1.019; 1.037)	1.033	(1.014; 1.051)
	M	1.037	(1.029; 1.046)	1.037	(1.020; 1.055)

Table 3: Estimated annual trends with 95%-confidence intervals for prevalence and incidence.

Treatment	Gender	Mortality		Relative Mortality	
All Antidiabetic Medications	F	.976	(.952; 1.001)	.983	(.958; 1.009)
	M	.966	(.943; .990)	.979	(.954; 1.004)
Insulin	F	.966	(.927; 1.007)	.973	(.933; 1.014)
	M	.971	(.932; 1.012)	.985	(.945; 1.027)
Oral Antidiabetic Medications	F	.979	(.950; 1.010)	.986	(.956; 1.017)
	M	.962	(.933; .990)	.974	(.945; 1.003)

Table 4: Estimated annual trends with 95%-confidence intervals for mortality.

9.2 Absolute numbers

Year	Size	Prevalence	Incidence	Mortality
1993	232,142	3,166	478	241
1994	232,732	3,261	471	216
1995	233,083	3,344	439	254
1996	233,286	3,374	548	211
1997	234,625	3,582	464	237
1998	234,814	3,651	465	231
1999	235,064	3,740	531	232

Table 5: Females, absolute numbers

Year	Size	Prevalence	Incidence	Mortality
1993	226,219	3,249	555	249
1994	226,912	3,415	567	264
1995	227,248	3,557	512	232
1996	227,482	3,673	618	235
1997	228,923	3,900	563	241
1998	229,214	4,073	628	244
1999	229,268	4,290	616	266

Table 6: Males, absolute numbers

10 Conclusion and Discussion

- Dynamic run-in avoids choosing length of run-in period
- Dynamic run-in should not be used for estimating secular trends
- Static run-in can be used for estimating secular trends
- Static run-in periods relies on constant misclassification rates
- Run-in periods induces selection bias
- More work is needed on methods avoiding run-in periods

10.1 Statistical take home messages

- Be specific about what is actually estimated
- A specific source of bias for one measure may not affect a related measure
- Look out for interpretation bias

10.2 Epidemiologic take home messages

- Do not always look for explanations in relative measures
- Do not confuse prevalence with incidence

References

- Hallas, J., D. Gaist, and L. Bjerrum (1997). The waiting time distribution as a graphical approach to epidemiologic measures of drug utilization. *Epidemiology* 8, 666–70.
- Mantel-Teeuwisse, A. K., O. H. Klungel, W. M. M. Verschuren, A. Porsius, and A. de Boer (2001). Comparison of different methods to estimate prevalence of drug use by using pharmacy records. *Journal of Clinical Epidemiology* 54(11), 1181–6.
- Rahimtoola, H., H. Buurma, C. C. Tijssen, H. G. Leufkens, and A. C. G. Egberts (2002). Incidence and determinants of migraine prophylactic medication in the netherlands. *Eur J Clin Pharmacol* 58(2), 149–55.
- Rahimtoola, H., H. Buurma, C. C. Tijssen, H. G. Leufkens, and A. C. G. Egberts (2003). Migraine prophylactic medication usage patterns in the netherlands. *Cephalalgia* 23(4), 293–293.

Støvring, H. (2002). *New Statistical Methods for Estimation of Prevalence, Incidence and Mortality based on Pharmacoepidemiological and other Health-related Databases*. Ph. D. thesis, Faculty of Health Sciences, University of Southern Denmark. <http://www.isd.sdu.dk/~stovring/thesis.pdf>.

Støvring, H., M. Andersen, H. Beck-Nielsen, A. Green, and W. Vach (2003). Rising prevalence of diabetes: evidence from a Danish pharmaco-epidemiological database. *The Lancet* 362, 537–38.

Støvring, H. and W. Vach (2004). Estimation of prevalence and incidence based on occurrence of health-related events. *Submitted*.