

## 순차패턴 마이닝 기법을 이용한 심방세동과 관련된 심혈관질환의 양상분석

계명대학교 의과대학 내과학교실, 의료정보학 교실<sup>1</sup>, 보건의료정보기술연구소<sup>2</sup>

조윤경 · 김윤년 · 김현태 · 김인철 · 김형섭 · 남창욱 · 한성욱 · 허승호 · 김권배 · 박희준<sup>1</sup> · 진종식<sup>2</sup>

### **Analysis of Coronary Vascular Diseases Associated with Atrial Fibrillation Using Sequential Mining**

Yun Kyeong Cho, M.D., Yoon Nyun Kim, M.D., Hyun Tae Kim, M.D., In Chul Kim, M.D.,  
Hyungseop Kim, M.D., Chang Wook Nam, M.D., Sung Wook Han, M.D.,  
Seung Ho Hur, M.D., Kwon Bae Kim, M.D., Hee Joon Park<sup>1</sup>, Jong Sik Jin<sup>2</sup>

*Department of Internal Medicine, Department of Medical Informatics<sup>1</sup>  
Keimyung University School of Medicine, Daegu, Korea  
Biomedical Informatics Technology Center<sup>2</sup>, Keimyung University Daegu, Korea*

**Abstract :** There are few reports about the incidence of cardiovascular diseases related to atrial fibrillation. Also, the nature of diseases is changing according to social environment. Korea is moving from a developing country into a developed country. Therefore, the nature of diseases is becoming the pattern of developed country. The aim of this study was to analyze the relationship between atrial fibrillation and cardiovascular diseases which mostly occur in developed country at this point. By sequential data mining, we analyzed the medial records, which were collected retrospectively, to evaluate the relationship between atrial fibrillation and cardiovascular diseases. The common diseases highly related with atrial fibrillation were hypertension (30%), heart failure (28%) and stroke (20%). The cardiovascular diseases preceding atrial fibrillation were hypertension (11%), heart failure (10%) and stroke (8%). The cardiovascular diseases following atrial fibrillation were hypertension (10%), heart failure (8%) and stroke (6%). In this study, we could understand the relationship between atrial fibrillation and cardiovascular diseases. Further study could find the causal relationship between atrial fibrillation and cardiovascular diseases more specifically.

**Key Words :** Atrial fibrillation , Mining, Sequential patterns

## 서론

질병 치료에 대한 개념은 이미 발병된 질환을 치료하는 방향으로 연구 및 개발되어 왔으나, 한편에서는 질병을 조기 발견하여 질병의 발현이나 진행을 중지 혹은 지연시키거나 질병을 예방하고자 하는 쪽으로도 동시에 진행되어 왔다. 그리고 최근에는 각개인의 유전적인 정보를 조사, 향후에 일어날 질병의 발생 경향을 예측하여 질병이 발생하였을 경우 가장 적절한 치료법을 적용하려고 하는 소위 맞춤의학이라는 새로운 개념도 나타나고 있다.

일반적으로 과거에는 질환의 양상이 지역별로 편중되어 있거나, 사회적 환경에 따라 혹은 경제적 상태나 변경에 의해서도 달라지는 것으로 보았다. 다시 말하면 사회적, 경제적, 혹은 정치적인 환경 변화에 따라 그 사회의 질환들은 발생양상이 달라졌다. 최근 우리의 환경은 경제적 여건이 급속히 좋아지면서 국민들의 수명이 길어지고 노인인구의 비율이 증가하여 고령화 사회 혹은 초 고령화 사회로 되어가고 있다고 한다. 이러한 경제적, 사회적, 정치적 환경이 바뀌면서 질환의 발생 현상도 달라지고 있다. 이러한 시점에서 우리사환의 일부이기는 하나 현재의 질병 양상을 좀더 객관적인 방법을 이용하여 분석하는 것은 의의가 있다고 하겠다.

이처럼 서구의 선진국들과 마찬가지로 우리나라에서도 경제의 눈부신 발전으로 국민의 영양상태가 좋아지고 질병의 진단기술 및 치료법이 발전되고 국가의 질병 예방 노력 등에 의해 국민의 수명이 연장되어 고령화 사회로 접어들었다. 고령화 인구가 많아지면 이에 따른 질병의 형태나 질병의 발생 양상이 달라지게 되기 때문에 고령과 연관된 질병의 예측이나 치료에 대한 많은 연구가 필요하다.

Framingham 연구[1]에 의하면 나이가 증가할수록 심방세동의 발생빈도는 증가하며 심방세동에 의한 합병증도 증가한다. 심방세동은 임상적으로는 심방수축이 없어지므로 심기능이 감소하여 심부전의 증상을 유발하고, 또한 심방 내 혈액의 저류로 응고인자가 증가하여 혈전의 발생이 많아져 전색증이 발생하게 된다. 또한 심방세동의 발생에는 연령뿐만 아니라 심장질환이나 고혈압 등 많은 인자들

이 영향을 주는 것으로 알려져 있다[2-5]. 하지만 국내에서는 심방세동의 의한 합병질환과 심방세동의 발생에 관련된 심혈관 질환들의 연관성에 대한 연구는 많지 않다.

이 연구는 심방세동과 관련하여 심방세동을 유발시킬 수 있는 심혈관계질환과 심방세동에 의해 발생하는 질환들의 연관성을 밝히고자 하였다.

## 연구대상 및 방법

### 연구대상

1996년 10월부터 2007년 10월까지 계명대학교 동산의료원에 내원한 환자들 중 심방세동을 가진 환자 3,826명을 대상으로 전자의무기록(EMR)에 기록된 관련 심혈관 질환의 병명, 즉 international classification of disease(ICD)의 I class를 조사하여 분석하였다. 의무기록에 기록된 질병의 발생시간을 분석하여 질병과 심방세동 사이의 선행유무를 판단하였다. 기존의 타 연구들에서 심방세동의 원인으로 알려져 있는 갑상선기능항진증과 같은 심혈관계 이외의 질환은 제외하였다.

### 방법

자료는 후향적으로 수집하여 순차패턴 마이닝 기법을 적용하여 분석하였다. 데이터 마이닝 기법 가운데 순차패턴 마이닝은 연관 규칙에 시간의 개념을 첨가하여 시간의 흐름에 따른 항목들의 상호 연관성을 찾는 것인데 즉, 순차패턴 마이닝 기법은 사용자가 지정한 최소지지도를 만족하는 항목들을 추출하고 이들 가운데 최대 빈도를 가지는 항목을 찾아내는 것이다(Formula 1).

이 연구에서 순차패턴 규칙  $X \Rightarrow Y$ 는 "질병 X가 발생하면 일정한 시간이 경과한 다음 질병 Y가 발생한다."라고 해석한다. 연관성 규칙과 순차패턴의 차이점은 연관성 규칙은  $X \Rightarrow Y$ ,  $Y \Rightarrow X$ 가 성립하지만 순차패턴은 시간의 개념이 추가되므로  $X \Rightarrow Y$ 만 성립한다. 이 연구에서는 순차패턴 마이닝을 적용하

$$\text{Support (S)} = \frac{\text{Number of disease X U Y}}{\text{Total number of disease}}$$

**Formula 1. Support of disease**

기 위하여 SAS사(Cary, USA)의 Enterprise Miner 4.1을 사용하였다.

연구방법은 크게 세 가지 단계로 나누어 실시하였다. 첫째, 전자의무기록에서 심혈관 질환을 가지고 있는 환자별로 상병 데이터를 추출하였다. 둘째, 추출된 데이터에 보다 정확한 순차패턴 마이닝 알고리즘을 적용하기 위하여 동일 질환명과 같은 case number를 가지는 데이터를 제거함으로써 중복을 배제하였다. 이러한 이유는 한 환자가 동일한 질환으로 여러 번 진단을 받게 되었을 때 순차패턴에서 동일한 질환명이 높은 지지도를 가지는 것을 해결하기 위한 것이다. 셋째, 최소지지도와 신뢰도를 만족하는 순차패턴을 찾아내고 규칙의 타당성 평가 및 분석을 하였다(Fig. 1).

**성 적**

**심방세동과 연관된 심혈관계 질환의 분포**

심혈관계 질환으로 심방세동과 연관성이 높은

질환으로는 고혈압(I10)이 1,158명(30%), 심부전(I50)이 1,987명(28%), 뇌혈관질환(I63)이 781명(20%), 허혈성 심질환(I20)이 581명(11%), 류마티스 승모판 질환(I05)이 421명(11%)의 순이었다(Table 1).

**심방세동 전 심혈관계 질환의 분포**

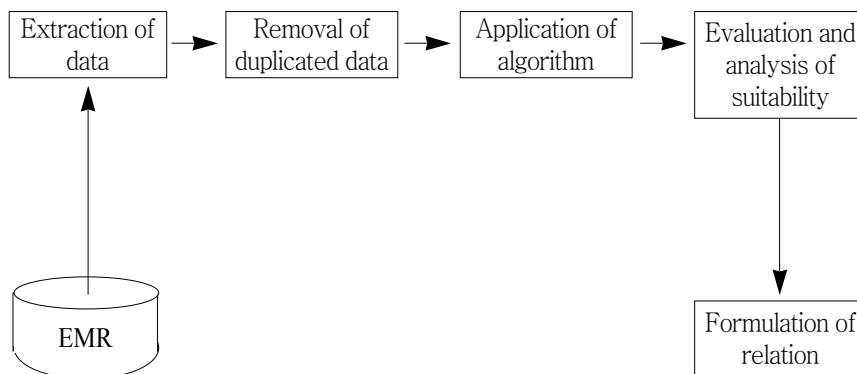
전자의무기록에 기록된 순서에 의해 심방세동 전에 발생한 심혈관계 질환은 고혈압이 407명(10.6%), 심부전이 383명(10.0%), 뇌혈관질환이 335명(8.8%), 허혈성 심질환이 315명(8.2%), 류마티스 승모판 질환이 190명(5.0%)의 순이었다(Table 2).

**심방세동 후 심혈관계 질환의 분포**

심방세동의 기록 후 발생한 심혈관계 질환의 분포는 고혈압이 371명(9.7%), 심부전이 306명(8%), 뇌혈관질환이 240명(6.3%), 허혈성 심질환이 154명(4.0%), 비류마티스 승모판 질환이 128명(3.4%)의 순이었다(Table 3).

**선행질환에 의한 심방세동 발생까지 경과시간의 분포**

선행질환으로 고혈압 진단후 심방세동이 발생



**Fig. 1.** The method of sequential mining. EMR: electronic medical record.

**Table 1.** Cardiovascular diseases related with atrial fibrillation (n=3,826)

ICD code	Disease	Number (support)
I10	Hypertension	1158(30)
I50	Heart failure	1087(28)
I63	Cerebral infarction	781(20)
I20	Angina pectoris	581(11)
I05	Rheumatic mitral valve disease	421(11)
I34	Non-rheumatic mitral valve disorder	335( 9)
I49	Other cardiac arrhythmias	256( 7)
I21	Acute myocardial infarction	244( 6)
I47	Paroxysmal tachycardia	182( 5)
I42	Cardiomegaly	174( 5)
I25	Chronic ischemic heart disease	167( 4)
I08	Multiple valve diseases	163( 4)
I69	Sequelae of cerebrovascular disease	115( 3)
I35	Non-rheumatic aortic valve disorder	104( 3)
I07	Rheumatic tricuspid valve disease	94( 2)
I11	Hypertensive heart disease	93( 2)
I70	Atherosclerosis	80( 2)
I15	Secondary hypertension	78( 2)
I74	Arterial embolism and thrombosis	67( 2)
I06	Rheumatic aortic valve disease	65( 2)
I65	Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction	59( 2)
I61	Intracranial hemorrhage	56( 1)
I45	Other conduction disorders	55( 1)
I44	Atrioventricular and left bundle branch block	53( 1)
I84	Hemorrhoids	51( 1)
I22	Subsequent myocardial infarction	48( 1)
I67	Other cerebrovascular diseases	47( 1)
I36	Non-rheumatic tricuspid valve disorder	46( 1)
I66	Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction	45( 1)
I71	Aortic aneurysm and dissection	43( 1)

Minimum support = 1

**Table 2.** Incidence of cardiovascular diseases previously diagnosed before atrial fibrillation (n=3,826)

History of disease	Number of transaction	Support (%)
I10 ⇒ I48	407	10.6
I50 ⇒ I48	383	10.0
I63 ⇒ I48	335	8.8
I20 ⇒ I48	315	8.2
I05 ⇒ I48	190	5.0
I21 ⇒ I48	114	3.0
I49 ⇒ I48	98	2.6
I34 ⇒ I48	88	2.3
I47 ⇒ I48	83	2.2
I42 ⇒ I48	68	1.8
I25 ⇒ I48	54	1.4
I08 ⇒ I48	49	1.3
I63 ⇒ I10 ⇒ I48	47	1.2
I15 ⇒ I48	46	1.2
I10 & I63 ⇒ I48	46	1.2
I11 ⇒ I48	42	1.1
I20 ⇒ I10 ⇒ I48	40	1.1
I70 ⇒ I48	39	1.0

Minimum support = 1

하는 기간은  $666 \pm 835$ 일, 심부전의 경우는  $634 \pm 867$ 일이었다. 뇌혈관질환 후 심방세동이 발생하는 기간은  $358 \pm 483$ 일이었으며 허혈성 심질환이 진단된 후 심방세동이 관찰되는 기간은  $829 \pm 912$ 일이었다. 또한 류마티스 승모판 질환 후 심방세동이 관찰되는 기간은  $1058 \pm 1060$ 일이었다. 심근경색 후에 심방세동의 관찰은  $367 \pm 653$ 일로 다빈도 질환 중에서 뇌졸중과 더불어 서서히 진행되는 고혈압, 심부전, 판막질환에 비해 비교적 짧은 시간에 발생하였다(Table 4).

**Table 3.** Incidence of the cardiovascular diseases happened after atrial fibrillation (n=3,826)

History of disease	Number of transaction	Support (%)
I48 ⇒ I10	371	9.7
I48 ⇒ I50	306	8.0
I48 ⇒ I63	240	6.3
I48 ⇒ I20	154	4.0
I48 ⇒ I34	128	3.4
I48 ⇒ I05	114	3.0
I48 ⇒ I49	94	2.5
I48 ⇒ I25	77	2.0
I48 ⇒ I69	75	2.0
I48 ⇒ I21	71	1.9
I48 ⇒ I08	66	1.7
I48 ⇒ I42	49	1.3
I48 ⇒ I47	43	1.1
I48 ⇒ I07	41	1.1
I48 ⇒ I35	41	1.1

Minimum support = 1

### 심방세동에 의한 후발질환 발생까지 경과시간의 분포

심방세동 진단 후 관찰되는 심혈관계 질환으로서 고혈압이  $482 \pm 720$ 일이었으며 심부전은  $380 \pm 578$ 일이었으며 뇌혈관질환은  $539 \pm 601$ 일이었고 허혈성 심질환의 발현은  $343 \pm 585$ 일이었다(Table 5).

## 고 찰

질환의 연관성의 연구나 역학적 연구는 다양한 형태로 진행되었다. 과거 종이의무기록으로 진행하

**Table 4.** Time interval from preceding cardiovascular disease to atrial fibrillation

History of disease	Average (day)	Standard deviation (day)	Minimum (day)	Maximum (day)
I10 ⇒ I48	666	835	1	3912
I50 ⇒ I48	634	867	1	4128
I63 ⇒ I48	358	483	1	2445
I20 ⇒ I48	829	912	1	3461
I05 ⇒ I48	1058	1060	2	3752
I21 ⇒ I48	367	653	1	3357
I49 ⇒ I48	628	751	1	3223
I34 ⇒ I48	680	755	14	2734
I47 ⇒ I48	624	1006	1	3333
I42 ⇒ I48	396	432	5	1616
I25 ⇒ I48	381	692	8	2984
I08 ⇒ I48	910	1117	8	2935
I05 ⇒ I48	1058	1060	2	3752
I70 ⇒ I48	552	449	3	1365
I15 ⇒ I48	565	639	14	2581
I11 ⇒ I48	913	793	87	2905

는 경우는 많은 시간과 전문 인력이 동원되었지만 그 결과에 의한 다양한 연관성을 밝히는 데는 많은 제한점이 있다. 최근에는 전자의무기록 및 컴퓨터 기술의 발전으로 대량의 정보를 저장하고 가공할 수 있게 되었다. 하지만 의료기관 내에서 진료정보의 컴퓨터화는 국외뿐 아니라 국내에서도 대단히 제한적으로 적용되고 있다. 국내에서는 10여년 전부터 전자의무기록을 사용하게 되었으며 또한 많은 전산화된 자료가 발생하게 되었다. 우선 본 연구는 전자의무기록에 저장된 진단자료를 이용하여 진행되었으며 이는 다른 연구와 차이를 보이고 있다.

연관성 분석 중 순차적 패턴분석은 시간에 따른 연관성을 보여줌으로 한 질환과 다음 질환과의 연

관성뿐만 아니라 질병의 발생 예측도 가능하다. 또한 질병의 연관성 분석에서 순차적 패턴을 적용한 연구는 국내에서는 많지 않으며[6] 다양한 분석방법을 적용하였다는데 본 연구의 의의가 있다.

질환의 유형이나 예측은 역학적 조사에서 많이 진행되어 왔으며 주로 위험인자를 발굴하고 이를 통해 질병의 유형을 파악하고 예측하려고 하였으며 최근에는 유전정보를 이용하여 발병가능성이 높은 객체를 발굴하고 질병을 예방하고자 하는 방향으로 진행되고 있다. 다양한 질환들이 존재하고 연구의 대상이 될 수 있지만 최근 의학발전과 경제의 발전, 그리고 환경의 개선 등으로 사람의 수명이 길어져 세계적으로 고령화 추세가 나타나며 이에 따른 심

**Table 5.** Time interval from atrial fibrillation to cardiovascular diseases

History of disease	Average (day)	Standard deviation (day)	Minimum (day)	Maximum (day)
I48 ⇒ I10	482	720	1	3772
I48 ⇒ I50	380	578	1	3711
I48 ⇒ I63	539	601	1	3090
I48 ⇒ I20	343	585	1	3250
I48 ⇒ I34	325	567	1	3286
I48 ⇒ I05	240	511	3	3343
I48 ⇒ I49	486	616	1	2127
I48 ⇒ I25	349	568	12	2519
I48 ⇒ I69	724	671	31	2162
I48 ⇒ I21	613	766	7	3157
I48 ⇒ I08	422	596	3	2251
I48 ⇒ I42	327	528	1	2079
I48 ⇒ I47	672	1134	5	3879
I48 ⇒ I07	335	429	1	1375
I48 ⇒ I35	357	623	3	1997

장질환이 주목을 받고 있다. 최근에는 심장질환 중 심방세동이 이슈화되고 있다.

연구결과에서 심방세동과 연관성이 있는 것으로 나타난 질환의 유형은 고혈압[7,8], 심부전[8,9], 뇌혈관질환[10,11], 허혈성 심질환[7-9,14-19], 그리고 심장 판막 질환[7,8,11-13] 순으로 다른 연구와 동일한 결과를 보였다. 그러나 기록된 순서에 의한 분석으로 볼 때 심방세동을 일으키는 원인 질환으로도 고혈압과 심부전, 뇌혈관 질환, 허혈성 심질환, 그리고 심장판막 질환 순으로 많았으며, 심방세동에 의해 발생할 것으로 보이는 질환도 같은 순서를 보였다.

## 요 약

이 연구는 선진국형의 질환인 심혈관계질환과 심방세동 간에 연관성이 있는지를 연구하고자 하였다. 후향적 방법으로 의무기록에 기록된 자료를 이용하여 심방세동과 타 심혈관계질환과의 시간적 연관성을 순차페턴 마이닝 기법을 사용하여 평가하였다. 심방세동과 가장 연관성이 많은 질환은 고혈압(30%), 심부전(28%), 뇌혈관질환(20%)이었다. 이중 심방세동 전에 진단된 질환은 고혈압(11%), 심부전(10%), 뇌혈관질환(8%) 순이었으며 심방세동 후 발견된 질환은 고혈압(10%), 심부전(8%), 뇌혈관질환(6%) 순이었다. 이 연구를 통하여 심방세동과 심혈관계 질환 상호간의 경향을 알

수 있으며 향후 세부 연구를 이용하여 좀더 구체적으로 심방세동과 심혈관계 질환과의 인과관계에 대해 연구하는 것이 필요하겠다.

## 참고 문헌

1. Kannel WB, Abbott RD, Savage DD, McNamara PM. Epidemiologic features of chronic atrial fibrillation: the Framingham study. *N Engl J Med* 1982;**306**:1018-22.
2. Majeed A, Moser K, Carroll K. Trends in the prevalence and management of atrial fibrillation in general practice in England and Wales, 1994-1998: analysis of data from the general practice research database. *Heart* 2001;**86**:284-8.
3. Feinberg WM, Blackshear JL, Laupacis A, Kronmal R, Hart RG. Prevalence, age distribution, and gender of patients with atrial fibrillation. Analysis and implications. *Arch Intern Med* 1995;**155**:469-73.
4. Heeringa J, van der Kulip DA, Hofman A, Kors JA, van Herpen G, Stricker BH, *et al.* Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. *Eur Heart J* 2006;**27**:949-53.
5. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, *et al.* Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the anticoagulation and risk factors in atrial fibrillation (ATRIA) study. *JAMA* 2001;**285**:2370-5.
6. 진중식, 박희준, 이정현, 김윤년, 윤경일, 엄홍섭. 순차패턴 마이닝을 이용한 상병의 연관성 분석. *한국경영정보학회지* 2007;**1**:614-8.
7. Psaty BM, Manolio TA, Kuller LH, Kronmal RA, Cushman M, Fried LP, *et al.* Incidence of and risk factors for atrial fibrillation in older adults. *Circulation* 1997;**96**:2455-61.
8. Krahn AD, Manfreda J, Tate RB, Mathewson FA, Cuddy TE. The natural history of atrial fibrillation: incidence, risk factors, and prognosis in the Manitoba follow-up study. *Am J Med* 1995;**98**:476-84.
9. Lloyd-Jones DM, Wang TJ, Leip EP, Larson MG, Levy D, Vasan RS, *et al.* Lifetime risk for development of atrial fibrillation: the Framingham heart study. *Circulation* 2004;**110**:1042-6.
10. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham study. *Stroke* 1991;**22**:983-8.
11. Ruigomez A, Garcia Rodriguez LA, Johansson S, Wallander MA, Edvardsson N. Risk of cerebrovascular accident after a first diagnosis of atrial fibrillation. *Clin Cardiol* 2007;**30**:624-8.
12. Kramer RJ, Zeldis SM, Hamby RI. Atrial fibrillation; a marker for abnormal left ventricular function in coronary heart disease. *Br Heart J* 1982;**47**:606-8.
13. Grigioni F, Avierinos JF, Ling LH, Scott CG, Bailey KR, Tajik AJ, *et al.* Atrial fibrillation complicating the course of degenerative mitral regurgitation: determinants and long-term outcome. *J Am Coll Cardiol* 2002;**40**:84-92.
14. Diker E, Aydogdu S, Ozdemir M, Kural T, Polat K, Cehreli S, *et al.* Prevalence and predictors of atrial fibrillation in rheumatic valvular heart disease. *Am J Cardiol* 1996;**77**:96-8.
15. Cameron A, Schwartz MJ, Kronmal RA, Kosinski AS. Prevalence and significance of atrial fibrillation in coronary artery disease (CASS Registry). *Am J Cardiol* 1988;**61**:714-7.
16. Liberthson RR, Salisbury KW, Hutter AM Jr, DeSanctis RW. Atrial tachyarrhythmias in acute myocardial infarction. *Am J Med* 1976;**60**:956-60.
17. Crenshaw BS, Ward SR, Granger CB, Stebbins AL, Topol EJ, Califf RM. Atrial fibrillation in the setting of acute myocardial infarction: the GUSTO-I experience. Global utilization of streptokinase and TPA for occluded coronary arteries. *J Am Coll Cardiol* 1997;**30**:406-13.
18. Wong CK, White HD, Wilcox RG, Criger DA, Califf RM, Topol EJ, *et al.* New atrial fibrillation after



acute myocardial infarction independently predicts death: the GUSTO-III experience. *Am Heart J* 2000;**140**:878-85.

19. Eldar M, Canetti M, Rotstein Z, Boyko V, Gottlieb S, Kaplinsky E, *et al.* Significance of paroxysmal atrial fibrillation complicating acute myocardial infarction in the thrombolytic era. SPRINT and thrombolytic survey groups. *Circulation* 1998;**97**:965-70.