

이중구간중도절단된 생존자료의 생존함수 비교를 위한 검정: 한국인 암 예방연구 중 당뇨병에의 응용

지선하¹ · 남정모² · 김진흠³

¹연세대학교 보건대학원 국민건강증진연구소, ²연세대학교 의과대학 예방의학교실,
³수원대학교 통계정보학과

(2009년 1월 접수, 2009년 3월 채택)

요약

이중구간중도절단된 자료의 생존함수를 비교하기 위한 두 검정법을 소개하고 한국인 암 예방연구 (Jee 등, 2005) 자료에 적용하여 당뇨병 잠복시간의 분포를 성별과 연령에 따라 비교하였다. Kim 등 (2006)의 검정법을 이중구간중도절단된 자료로 확장한 검정법은 위험집합의 크기에만 의존하는 가중값을 사용하기 때문에 대용량 자료의 분석에서 Sun (2006)의 검정법보다 계산 시간을 대폭 줄일 수 있으며, 이산형 생존자료뿐만 아니라 연속형 생존자료에도 적용 가능한 장점이 있다. 당뇨병의 잠복시간이 성별에 따라 매우 유의하게 달랐으며 여자의 잠복시간이 남자보다 긴 것으로 나타났다. 4개 연령그룹 간 당뇨병의 잠복시간도 성별에 관계 없이 매우 유의하게 달랐으며 여자의 경우가 남자의 경우보다 그 차이가 훨씬 더 유의했다. 한편, 소표본 모의실험을 통해 제안한 검정법과 Sun (2006)의 검정법의 검정력을 비교하였으며 제안한 검정법의 검정력이 Sun (2006)의 검정법보다 더 좋은 것으로 나타났다.

주요용어: 이중구간중도절단, 생존함수, 로그-순위 검정, 다중대체, 잠복시간.

1. 서론

임상 시험 및 역학 연구에서 주요 과제 중 하나는 서로 다른 처리 효과의 생존함수를 비교하는 것이다. 본 논문에서는 수집된 자료가 이중구간중도절단(doubly interval-censored)되었을 때 $p(\geq 2)$ 개 처리 그룹의 생존함수를 비교하기 위한 검정법을 다루고자 한다.

관심 있는 생존시간이 연관된 두 사건, 즉 시작사건과 사망사건의 발생 간격으로 정의되고, 이 두 사건의 발생시점이 구간중도절단된(interval-censored) 자료를 이중구간중도절단된 자료라고 부른다. 만일 시작사건이 정확히 관측되면 이 자료는 구간중도절단된 자료가 되고, 만일 시작사건이 정확히 관측되고 동시에 사망사건이 우중도절단되면(right-censored) 이때는 우중도절단된 자료가 된다. 이중구간중도절단된 자료의 대표적인 예는, HIV(human immunodeficiency virus)에 이미 감염되었거나 HIV에 감염될 위험에 있는 환자들이 AIDS(acquired immune deficiency syndrome)로 진행될 위험에 노출되어 있는 추적연구에서 찾아 볼 수 있다. 이 예에서 생존시간은 HIV 감염시점과 AIDS 발병시점 사이의 간격으로 정의되며 흔히 이를 AIDS 잠복시간(incubation time)이라고 한다. 그러나 AIDS 연구에서는 사

이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2007-314-C00042).

³교신저자: (445-743) 경기도 화성시 봉담읍 와우리 산 2-2, 수원대학교 통계정보학과, 교수.

E-mail: jinhkim@suwon.ac.kr

전에 정한 시점에서만 주기적으로 모니터링 하기 때문에 HIV 감염시점이나 AIDS 발현시점이 종종 구간중도절단된 채로 관측되곤 한다.

서구국가들은 물론 최근 급격히 개발 중인 아시아에서도 당뇨병은 유행으로 번질 정도로 빠르게 증가하고 있다. 2006년 사망원인 통계자료에 따르면 당뇨병으로 인한 사망률은 23.7%로 악성신생물, 뇌혈관 질환, 심장질환에 이어 4위를 차지하고 있다 (통계청, 2007). 특히 제2형 당뇨병 환자의 경우 당뇨병이 임상적으로 진단되기 적어도 4년-7년 이전에 발병되며 (김진화 등, 2008), 당뇨병성 망막병증, 절단에 이르는 당뇨병성 신경병증, 당뇨병성 신증 등 미세혈관 합병증이 동반되기도 하기 때문에 (Fonseca, 2007) 제2형 당뇨병의 자연사(natural history)를 이해하는 것은 중요하다. 제3기 국민건강영양조사 보고서 (Korea Centers for Disease Control and Prevention, 2007)에 따르면 우리나라 30세 이상 연령군의 당뇨병 유병률은 약 9.1%이며, 17.4%(남자 21.3%, 여자 13.7%)가 당뇨병전증(prediabetes)을 가지고 있다. 내당능장애(impaired glucose tolerance: IGT)와 공복혈당장애(impaired fasting glucose: IFG)는 혈당치가 조금 높아진 것을 의미하며, IGT와 IFG는 흔히 당뇨병전증이라는 용어로 사용하고 있다. 당뇨병전증은 당뇨병 및 심혈관질환, 대사증후군 등 질병의 위험을 증가시킨다. 당뇨병전증 중에서 연 3%-10% 정도가 제2형 당뇨병으로 진행되며, IGT와 IFG를 모두 갖고 있는 사람은 IGT와 IFG 중에서 한 가지만 갖고 있는 사람보다 당뇨병 위험도가 훨씬 높으며 특히 정상혈당군보다 무려 6배나 당뇨병 위험이 증가한다. 그러나 당뇨병전증에서 당뇨병으로의 진행은 연구대상자의 인종, 연령, 성별, 혈당검사방법 및 판단기준에 따라 많은 차이를 보인다 (박이병, 2006). 본 논문에서는 당뇨병전증이 나타난 후 당뇨병으로 진단될 때까지의 잠복시간이 성별과 연령에 따라 서로 다른지를 통계적으로 검증하고자 한다. 그러나 검진이 주기적으로 이루어지기 때문에 정상혈당(diabetes-free)에서 당뇨병전증으로 진행된 시점을 정확히 알 수 없고 단지 정상으로 판정한 가장 최근 검진시점과 당뇨병전증으로 진단한 검진시점만을 알 수 있다. 마찬가지로 이유로 당뇨병전증에서 당뇨병으로 진행된 시점도 정확히 알 수 없기 때문에 본 논문에서 관심 있는 잠복시간의 양쪽 끝 값이 모두 구간중도절단될 수도 있다.

구간중도절단된 자료의 경우, 생존함수의 비교에 대한 연구가 매우 많다. 그 중에서 순위에 기초한 비모수적 검정법들을 중심으로 열거하면 다음과 같다. Self와 Grossman (1986), Fay (1996), Sun (1996), Pan (2000), Zhao와 Sun (2004), Kim 등 (2006). Zhao와 Sun (2004)은 우중도절단된 자료에 가장 널리 사용되는 로그-순위 검정법(log-rank test)을 구간중도절단된 자료로 일반화한 비모수적 검정법을 제안했다. Zhao와 Sun (2004)의 검정법은 Sun (1996)이 제안한 검정법보다 다음 두 가지 측면에서 볼 때 더 우수하다고 생각한다. 첫째, Zhao와 Sun (2004)의 검정법은 이산형 생존자료 뿐만 아니라 연속형 생존자료에도 적용 가능하고, 둘째, Zhao와 Sun (2004)의 검정법은 우중도절단된 자료에 대해 로그-순위 검정법과 정확하게 같다. Kim 등 (2006)은 EM(expectation-maximization) 알고리즘이 필요하지 않은 로그-순위형 검정법을 제안했는데, 이것은 추정 생존함수를 포함하는 가중값 대신에 각 관찰 시점에서 계산된 위험집합(risk set)의 크기에만 의존하는 균등 가중값(uniform weights)을 사용하기 때문에 가능하다.

이중구간중도절단된 자료의 경우, Sun (2001)은 이산형 생존자료에 적용 가능한 비모수적 검정법을 제안했다. 그러나 Sun (2001)의 검정법은 우중도절단된 자료에 대해 로그-순위 검정법으로 축소되지 않는 문제점을 갖고 있다. 이런 문제를 해결하기 위해 그는 자신의 검정법을 수정했는데 (Sun, 2006), 수정된 검정법은 Zhao와 Sun (2004)의 검정법을 이중구간중도절단된 자료로 확장한 것이다.

2절에서는 이중구간중도절단된 자료의 생존함수를 비교하기 위한 Sun (2006)의 비모수적 검정법을 소개하고, 또한 Kim 등 (2006)의 검정법을 이중구간중도절단된 자료로 확장하고자 한다. 3절에서는 2절에서 소개한 검정법들을 한국인 암 예방연구(Korean Cancer Prevention Study: KCPS; Jee 등, 2005) 자료에 적용하고 그 결과를 고찰하고자 한다. 4절에서는 소표본 모의실험을 통해 제안한 검정법과 Sun

(2006)의 검정법이 명목 유의수준을 잘 조절하는지 여부를 살펴보고자 하며, 또한 두 검정법의 검정력을 서로 비교하고자 한다. 마지막으로 5절에서는 본 논문의 한계점과 향후 연구과제를 살펴보고자 한다.

2. 일반화 로그-순위형 검정법

2.1. 기호와 가정

서로 다른 $p(\geq 2)$ 개의 처리 그룹에서 수집된 n 명의 개체로 이루어진 생존자료가 있다고 하자. X_i 와 S_i 는 각각 질병의 감염시점과 발생시점이고, $T_i = S_i - X_i$ 는 관심 있는 생존시간이다. 단, $i = 1, \dots, n$ 이고, X_i, S_i, T_i 는 모두 연속이라고 가정한다. 따라서 각 개체의 관찰값은 (L_i, R_i, U_i, V_i) 와 같은 형태로 주어지며, $L_i \leq X_i \leq R_i, U_i \leq S_i \leq V_i$ 이다. 본 논문에서는 서로 다른 $p(\geq 2)$ 개 처리 그룹의 생존시간의 분포가 동일한지를 검정하고자 한다. $S_q(t) = \Pr(T > t), q = 1, \dots, p$ 를 처리 그룹 q 의 생존함수라고 놓으면, 관심 있는 귀무가설 H_0 는 ‘모든 $t > 0$ 에 대해 $S_1(t) = \dots = S_p(t)$ 이다’라고 표현할 수 있다. 한편, X_i 와 T_i 는 서로 독립이고, $[L_i, R_i], [U_i, V_i]$ 는 각각 X_i, S_i 와 서로 독립이라고 가정한다.

2.2. Sun (2006)의 검정법

확률변수 X_i 와 T_i 가 0보다 큰 확률을 갖는 점들을 각각 $u_1 < \dots < u_r$ 와 $v_1 < \dots < v_{s-1} < v_s = \infty$ 라고 하자. $w_j = \Pr(X_i = u_j), j = 1, \dots, r$ 과 $f_k = \Pr(T_i = v_k), k = 1, \dots, s$ 로 정의하자. H_0 하에서 이들 w_j 와 f_k 의 최대우도추정량을 각각 \hat{w}_j 와 \hat{f}_k 로 표시하자. $w = \{w_j\}$ 와 $f = \{f_k\}$ 의 추정에 사용되는 대표적인 방법으로 De Gruttola과 Lagakos (1989)의 자기일치(self-consistency) 알고리즘을 이용하는 방법과 Gómez과 Lagakos (1994)의 2단계 추정법, Sun (1997)이 제안한 조건부우도함수(conditional likelihood)에 기초한 방법이 있다. 이들 중에서 처음 두 방법은 구간중도절단만을 가진 자료에 적용 가능하지만 마지막 방법은 구간중도절단 뿐만 아니라 절단(truncated)을 포함하는 자료에도 적용 가능하다. Gómez과 Lagakos (1994)는 주변우도함수(marginal likelihood)에 기초하여 w 를 추정한 후 f 를 주변적으로 추정한 반면, De Gruttola과 Lagakos (1989)는 전체우도함수(full likelihood)로부터 w 와 f 를 동시에 추정했다. $\alpha_{ik} = I(v_k \in [U_i - R_i, V_i - L_i])$ 과 $\hat{\alpha}_{ik} = \sum_{u_j + v_k \in [U_i, V_i]} \hat{w}_j$ 를 정의하고, 또한 $\delta_i = I(V_i - L_i < v_s)$ 과 $\rho_{ik} = I(\delta_i = 0, U_i - R_i \geq v_k)$ 를 정의하자. 따라서 δ_i 가 1이면 생존시간이 구간중도절단되었거나 정확히 관측된 개체이고, 0이면 우중도절단된 개체이다.

Sun (2006)은 $v_k, k = 1, \dots, s$ 시점에서 사망자 수 d_k 와 위험집합의 크기 n_k 를 다음과 같이 추정했다.

$$d_k = \sum_{i=1}^n \delta_i \frac{\alpha_{ik} \hat{\alpha}_{ik} \hat{f}_k}{\sum_{m=1}^s \alpha_{im} \hat{\alpha}_{im} \hat{f}_m}, \quad n_k = \sum_{i=1}^n \delta_i \frac{\sum_{j=k}^s \alpha_{ij} \hat{\alpha}_{ij} \hat{f}_j}{\sum_{m=1}^s \alpha_{im} \hat{\alpha}_{im} \hat{f}_m} + \sum_{i=1}^n \rho_{ik}.$$

식 n_k 의 첫 번째 항은 생존시간이 구간중도절단되었거나 정확히 관측된 개체가 위험집합에 기여하는 양을 의미하며, 그 크기는 v_k 시점까지의 상위 누적 조건부 확률값과 같다. 아울러 식 n_k 의 두 번째 항은 생존시간이 우중도절단된 개체가 위험집합에 기여하는 양을 의미하는데 중도절단이 일어난 시점까지는 1의 값을 갖고 그 이후로는 0의 값을 갖는다. 한편, 처리 그룹에 따른 추정 사망자수 d_{kq} 와 위험집합의 크기 n_{kq} 는 위 식 d_k 와 n_k 에서 해당 처리 그룹에 속하는 개체들만 합한 값으로 정의한다. 따라서 처리 그룹 q 의 로그-순위형 통계량 U_q 는 $U_q = \sum_{k=1}^{s-1} (d_{kq} - d_k n_{kq} / n_k)$ 로 주어진다.

가설 H_0 를 검정하기 위해 Sun (2006)은 $\mathbf{U} = (U_1, \dots, U_{p-1})'$ 에 기초하여 통계량 $S = \mathbf{U}' \hat{\mathbf{V}}^{-1} \mathbf{U}$ 를 제안했다. 여기서 $\hat{\mathbf{V}}$ 은 \mathbf{U} 의 공분산행렬에 대한 추정량이다. $\hat{\mathbf{V}}$ 으로 아래에서 소개할 추정량을 사용하

면, H_0 하에서 검정통계량 S 가 근사적으로 자유도가 $(p - 1)$ 인 카이제곱 분포를 따르기 때문에 $S > \chi_{p-1}^2(\alpha)$ 이면 유의수준 $\alpha \in (0, 1)$ 에서 H_0 를 기각한다 (Little과 Rubin, 2002).

한편, Sun (2006)은 \hat{V} 을 얻기 위한 한 방법으로 다중대체(multiple imputation) 방법을 제안했다. 이 중증도절단된 자료의 다중대체는 두 단계로 이루어진다. 첫 번째 단계에서는 구간중증도절단된 X_i 를 대체하고(\tilde{X}_i : 대체값), 두 번째 단계에서는 대체된 \tilde{X}_i 를 써서 구간중증도절단된 T_i 를 대체한다(\tilde{T}_i : 대체값). 여기서 \tilde{X}_i 는 아래 조건부 분포로부터 생성하고

$$\Pr(\tilde{X}_i = u_j | X_i \in [L_i, R_i]) = \frac{\hat{w}_j}{\sum_{m: u_m \in [L_i, R_i]} \hat{w}_m}, \quad u_j \in [L_i, R_i],$$

\tilde{T}_i 는 아래 조건부 분포로부터 생성한다.

$$\Pr(\tilde{T}_i = v_k | \tilde{X}_i, S_i \in [U_i, V_i]) = \frac{\hat{f}_k}{\sum_{m: v_m \in [U_i - \tilde{X}_i, V_i - \tilde{X}_i]} \hat{f}_m}, \quad v_k \in [U_i - \tilde{X}_i, V_i - \tilde{X}_i].$$

다중대체 후에는 우중증도절단된 자료로부터 \mathbf{U} 를 구하고 ($\mathbf{U}^{(b)}$), \mathbf{U} 의 공분산행렬에 대한 추정값을 구한다 ($\hat{\mathbf{V}}^{(b)}$) (Klein과 Moeschberger, 2003). 위 다중대체 과정을 $B(> 0)$ 번 반복하여 얻은 $\{(\mathbf{U}^{(b)}, \hat{\mathbf{V}}^{(b)})\}$, $b = 1, \dots, B$ 로 부터 $\hat{\mathbf{V}} = \hat{\mathbf{V}}_1 + (1 + B^{-1})\hat{\mathbf{V}}_2$ 를 구한다. 여기서 $\hat{\mathbf{V}}_1$ 는 $\hat{\mathbf{V}}^{(b)}$ 들의 평균이고, $\hat{\mathbf{V}}_2$ 는 $\mathbf{U}^{(b)}$ 들의 표본공분산행렬이다.

2.3. Kim 등 (2006)의 검정법의 확장

Kim 등 (2006)은 구간중증도절단된 자료의 생존함수 비교를 위해 EM 알고리즘이 필요하지 않은 로그-순위 검정법을 제안했다. 본 논문에서는 Kim 등 (2006)의 방법을 이중구간중증도절단된 자료로 확장하고자 한다.

$u_1^* < \dots < u_{s^*}^*$ 는 $\{L_i, R_i, i = 1, \dots, n\}$ 의 순서화 된 서로 다른 값이고, $v_1^* < \dots < v_{s^*-1}^* < v_{s^*}^* = \infty$ 는 $\{U_i - R_i, V_i - L_i, i = 1, \dots, n\}$ 의 순서화 된 서로 다른 값이다. 만일 $v_k^* \in [U_i - R_i, V_i - L_i]$, $k = 1, \dots, s^* - 1$ 이면 $\alpha_{ik}^* = \sum_{j=1}^{s^*} I(u_j^* + v_k^* \in [U_i, V_i], u_j^* \in [L_i, R_i])$ 로 정의하고, 만일 $v_k^* \notin [U_i - R_i, V_i - L_i]$ 이면 $\alpha_{ik}^* = 0$ 으로 정의한다. T_i 가 구간중증도절단되거나 정확하게 관측되면 $\delta_i^* = 1$ 로 정의하고, 우중증도절단되면 0으로 정의한다. H_0 하에서 개체 i 는 (X_i, T_i) 의 값으로 집합 $\mathcal{A}_i = \{(u_j^*, v_k^*); u_j^* + v_k^* \in [U_i, V_i], u_j^* \in [L_i, R_i], v_k^* \in [U_i - R_i, V_i - L_i]\}$ 위에 있는 값을 $1/\sum_{k=1}^{s^*-1} \alpha_{ik}^*$ 의 확률로 균등하게 갖는다고 가정하자.

균등모형 하에서 v_k^* , $k = 1, \dots, s^* - 1$ 시점에서의 사망자 수와 위험집합의 크기는 각각 다음과 같이 추정된다.

$$d_k^* = \sum_{i=1}^n \delta_i^* \frac{\alpha_{ik}^*}{\sum_{m=1}^{s^*-1} \alpha_{im}^*}, \quad n_k^* = \sum_{i=1}^n \frac{\sum_{l=k}^{s^*-1} \alpha_{il}^*}{\sum_{m=1}^{s^*-1} \alpha_{im}^*}.$$

처리 그룹에 따른 사망자 수 d_{kq}^* 와 위험집합의 크기 n_{kq}^* 는 위 식 d_k^* 와 n_k^* 에서 해당 처리 그룹에 속하는 개체들만 합한 값으로 정의한다. d_k^* , n_k^* , d_{kq}^* , n_{kq}^* 는 모두 v_k^* 에서 위험집합의 크기에만 의존하기 때문에 Sun (2006)과 다르게 X 와 T 에 대한 생존함수의 추정이 필요하지 않다. 이런 이유로 계산 시간을 줄일 수 있고, 또한 이산형 생존자료 뿐만 아니라 연속형 생존자료에도 적용 가능한 장점이 있다. 2.2절에서처럼 처리 그룹 q 의 로그-순위형 통계량 U_q^* 는 $U_q^* = \sum_{k=1}^{s^*-1} (d_{kq}^* - d_k^* n_{kq}^* / n_k^*)$ 로 주어지는데, 특히

X_i 가 정확히 관측되고 동시에 S_i 가 우중도절단되면 U_q^* 은 통상적인 로그-순위 통계량과 정확하게 일치한다 (Klein과 Moeschberger, 2003).

$\mathbf{U}^* = (U_1^*, \dots, U_{p-1}^*)'$ 라고 하자. 가설 H_0 를 검정하기 위해 검정통계량 $P = \mathbf{U}^{*'} \hat{\mathbf{V}}^{*-1} \mathbf{U}^*$ 을 제안한다. 여기서 $\hat{\mathbf{V}}^*$ 은 \mathbf{U}^* 의 공분산행렬에 대한 추정량이다. $\hat{\mathbf{V}}^*$ 으로 아래에서 소개할 추정량을 사용하면, H_0 하에서 P 는 근사적으로 자유도가 $(p - 1)$ 인 카이제곱분포를 따르기 때문에 $P > \chi_{p-1}^2(\alpha)$ 이면 유의수준 α 에서 H_0 를 기각한다 (Little과 Rubin, 2002).

한편, $\hat{\mathbf{V}}^*$ 은 Kim 등 (2006)의 방법과 Sun (2006)의 다중대체 방법을 써서 얻을 수 있다. 먼저 구간중도절단된 X_i 의 대체값 X_i^* 를 아래 조건부 균등분포로부터 생성한다.

$$\Pr(X_i^* = u_j^* | X_i \in [L_i, R_i]) = \frac{1}{\sum_{m=1}^{r^*} I(u_m^* \in [L_i, R_i])}, \quad u_j^* \in [L_i, R_i].$$

X_i^* 가 주어졌을 때 $\hat{v}_1 < \dots < \hat{v}_{s'-1} < \hat{v}_{s'} = \infty$ 는 $\{U_i - X_i^*, V_i - X_i^*, i = 1, \dots, n\}$ 의 순서화된 서로 다른 값이라고 하자. 구간중도절단된 T_i 의 대체값 T_i^* 를 아래 조건부 균등분포로부터 생성한다.

$$\Pr(T_i^* = \hat{v}_k | X_i^*, S_i \in [U_i, V_i]) = \frac{1}{\sum_{m=1}^{s'-1} I(\hat{v}_m \in [U_i - X_i^*, V_i - X_i^*])}, \quad \hat{v}_k \in [U_i - X_i^*, V_i - X_i^*].$$

이 때 우중도절단된 T_i 는 $T_i^* = U_i - X_i^*$ 로 대체한다. 위 두 단계의 대체 과정을 거쳐 우중도절단된 생존시간 자료 $\mathcal{D} = \{(T_i^*, \delta_i^*), i = 1, \dots, n\}$ 를 생성한다. 이제 우중도절단된 자료 \mathcal{D} 로부터 \mathbf{U}^* 를 구하고 이를 $\mathbf{U}^{*(b)}$ 로 놓자. 또한 \mathbf{U}^* 의 공분산행렬에 대한 추정값을 $\hat{\mathbf{V}}^{*(b)}$ 로 놓자 (Klein과 Moeschberger, 2003). 위와 같은 대체 과정을 $B(> 0)$ 번 반복하여 얻은 $\{(\mathbf{U}^{*(b)}, \hat{\mathbf{V}}^{*(b)}), b = 1, \dots, B\}$ 로부터 $\hat{\mathbf{V}}^* = \hat{\mathbf{V}}_1^* + (1 + B^{-1})\hat{\mathbf{V}}_2^*$ 를 구한다. 여기서 $\hat{\mathbf{V}}_1^*$ 은 $\hat{\mathbf{V}}^{*(b)}$ 들의 평균이고, $\hat{\mathbf{V}}_2^*$ 은 $\mathbf{U}^{*(b)}$ 들의 표본공분산행렬이다. 따라서 \mathbf{U}^* 의 공분산행렬에 대한 추정량으로 위 다중대체 방법으로 얻어진 $\hat{\mathbf{V}}^*$ 을 사용하면 통계량 P 는 우중도절단된 자료에 대한 통상적인 로그-순위 통계량과 정확하게 일치한다.

3. 적용 예

본 논문에서 분석하고자 하는 자료는 한국인 암 예방연구(KCPS)의 일부로서 연구설계에 대한 자세한 내용은 Jee 등 (2005)에서 찾아볼 수 있다. KCPS 코호트에서 1992년 건강검진자료가 조사된 피보험자 912,567명 중에서 아래 조건에 해당되는 사람들은 일차적으로 분석에서 제외했다. 단, 괄호 안에 있는 숫자는 해당 조건을 만족하는 사람수를 의미한다. 아래 5개의 조건 중에서 조건(3)을 만족하는 사람들은 연구시작 전에 이미 당뇨병전증을 갖고 있어 당뇨병 이환시간의 시작시점을 알 수 없기 때문에 제외했고(당뇨병전증이 좌중도절단된(left-censored) 경우에 해당), 조건(4)를 만족하는 사람들은 연구종료 전까지 당뇨병전증이 나타나지 않아 당뇨병 이환시간을 정의할 수 없기 때문에 제외했다(당뇨병전증이 우중도절단된 경우에 해당). 조건(5)를 만족하는 사람들은 시작시점에 대한 정보가 없어 잠복시간을 정의할 수 없기 때문에 제외했다. 한편, 본 논문에서는 당뇨병전증에서 당뇨병으로 진행되는 질병모델을 비가역적인 모델로 가정하였다. 다시 말해 어떤 시점에서 당뇨병전증으로 진행된 대상자가 다음 검진에서 정상혈당치로 관측되었다하더라도 당뇨병전증이 나타난 시점부터 당뇨병으로 진행하기 전까지의 관측값을 당뇨병전증으로 간주하고 분석하였다.

1. 연령이 20세 미만인 사람 (2,076명)
2. 1992년 시작시점에서 암을 비롯한 만성질환을 가진 사람 (67,221명)

표 3.1. 성별에 따른 연령별(G_1 : 20-29세, G_2 : 30-39세, G_3 : 40-49세, G_4 : 50세 이상) 표본수(n)와 우중도절단 비율(c_f)

연령그룹	남자		여자	
	n	c_f	n	c_f
G_1	5,107	0.931	7,955	0.980
G_2	28,495	0.904	23,473	0.968
G_3	20,329	0.872	10,243	0.943
G_4	13,407	0.890	4,213	0.927
합계	67,338	0.894	45,884	0.961

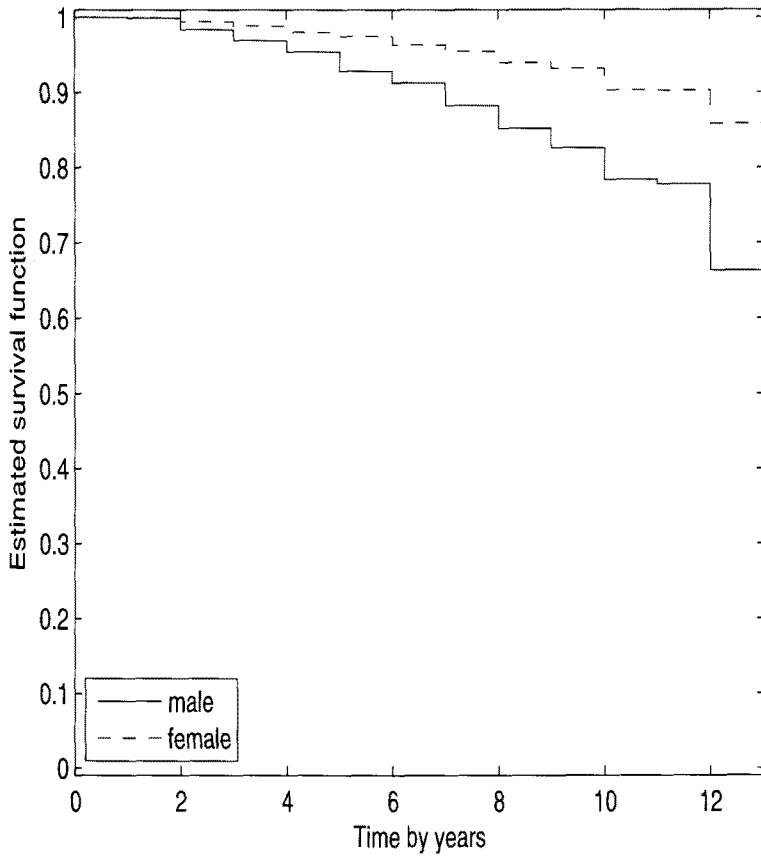


그림 3.1. 성별에 따른 당뇨병 잠복시간의 생존함수에 대한 최대우도추정량

3. 1992년 시작시점에서 공복시 혈당이 100mg/dL 이상인 당뇨병전증을 가진 사람 (160,571명)
4. 1992년부터 2005년까지 연구기간 동안 당뇨병전증으로 진행된 적이 없는 사람 (344,081명)
5. 정상혈당에서 당뇨병으로 바로 진행된 사람 (23,391명)

일차적으로 뽑힌 315,227명 중에서 남자는 269,343명 이었고, 여자는 45,884명 이었다. 그런데 남자의 경우는 표본의 크기가 너무 커서 De Gruttoda과 Lagakos (1989)의 생존함수 추정량이 수렴하지 못하

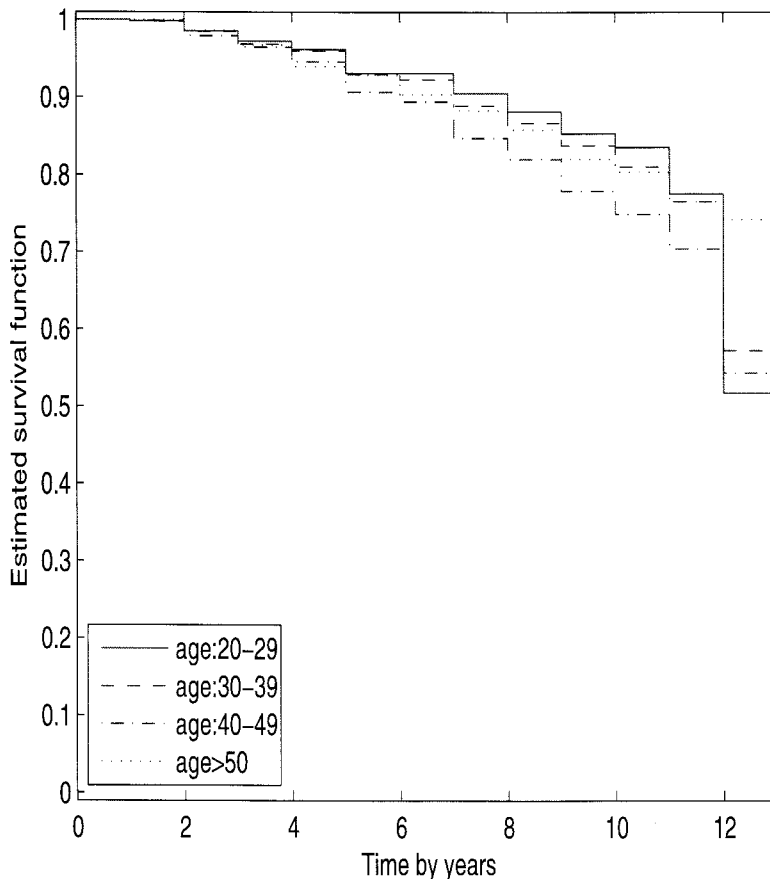


그림 3.2. 남자들의 4개 연령그룹('20-29세', '30-39세', '40-49세', '50세 이상')에 따른 당뇨병 잠복시간의 생존함수에 대한 최대우도추정량

는 문제가 발생하였다. 이를 해결하기 위해 남자의 경우는 약 25%를 임의추출하여 최종분석대상자로 선정하였는데, 추출률을 25%로 잡은 것은 De Gruttola와 Lagakos (1989)의 생존함수 추정이 물리적으로 가능한 최대 표본의 크기가 되었기 때문이다. 한편, 대표성 있는 표본을 얻기 위해 남자들을 연령과 당뇨병의 발현 유무(즉, 당뇨병의 우중도절단 여부)의 조합에 따라 8개의 층(stratum)으로 나눠 각 층에서 25%의 표본을 임의추출하였다. 이 때 연령은 '20-29세', '30-39세', '40-49세', '50세 이상'과 같이 4개 그룹으로 나누었다. 표 3.1은 성별에 따른 연령그룹별 표본수(n)와 우중도절단 비율(c_f)을 나타낸 것이다.

본 논문에서는 당뇨병진증부터 당뇨병 진단을 받을 때까지의 잠복시간에 대해 성별에 따라 연령그룹별 비교를 하였다. 그림 3.1은 성별에 따라 잠복시간의 생존함수를 추정한 것이다 (De Gruttola와 Lagakos, 1989). 그림 3.1로 볼 때 여자의 잠복시간이 남자의 잠복시간보다 훨씬 길다고 생각된다. 그림 3.2와 3.3은 각각 남자와 여자에 대해 연령그룹별로 잠복시간의 생존함수를 추정한 것이다 (De Gruttola와 Lagakos, 1989). 남자는(그림 3.2) '20-29세' 그룹의 잠복시간이 가장 길고 '40-49세' 그룹의 잠복시간이 가장 짧은 것으로 나타났으며, '30-39세' 그룹의 잠복시간이 '50세 이상' 그룹의 잠복시간보

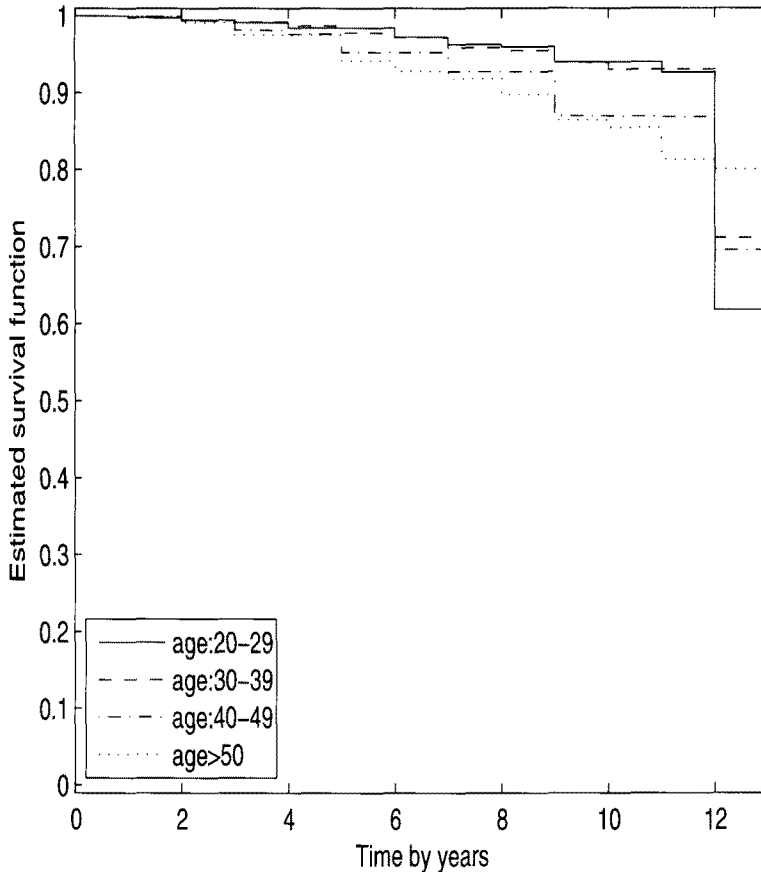


그림 3.3. 여자들의 4개 연령그룹('20-29세', '30-39세', '40-49세', '50세 이상')에 따른 당뇨병 잠복시간의 생존함수에 대한 최대우도추정량

다 약간 긴 것으로 나타났다. 여자도(그림 3.3) 남자와 마찬가지로 '20-29세' 그룹의 잠복시간이 가장 긴 반면 잠복시간이 가장 짧은 그룹은 '50세 이상' 그룹으로 나타났다. 나머지 두 그룹의 비교에서는 '30-39세' 그룹의 잠복시간이 '40-49세' 그룹의 잠복시간보다 훨씬 긴 것으로 나타났다.

한편, 연령그룹에 따라 잠복시간의 분포가 서로 다른지를 제안한 통계량 P 와 Sun (2006)의 통계량 S 를 써서 검정하였다. 이에 앞서 성별에 따라 잠복시간이 서로 다른지를 살펴보았는데, 두 검정통계량의 값이 각각 $P = 1345.53$ (유의확률 값 < 0.001)과 $S = 1238.64$ (유의확률 값 < 0.001)로 나와 그림 3.1에서 예상했던 것처럼 그 차이가 매우 유의했다. 성별에 따라 생존함수가 매우 달라서 이후의 분석에서는 성별에 따라 나누어 연령그룹 간 생존함수를 비교하였다. 표 3.2는 성별에 따라 제안한 통계량 P 와 Sun (2006)의 통계량 S 의 값과 $B = 50$ 번의 반복 대체에 의해 얻어진 유의확률 값을 나타낸 것이다. 4개 연령그룹을 동시에 비교하기에 앞서 두 그룹씩 짝지은 비교를 수행하였다. 두 그룹 간 비교는 모두 6 가지이며 각각 $G_1 vs. G_2$, $G_1 vs. G_3$, $G_1 vs. G_4$, $G_2 vs. G_3$, $G_2 vs. G_4$, $G_3 vs. G_4$ 이다. 성별에 관계 없이 6개 모든 비교가 다 유의했으며, $G_1 vs. G_2$ 의 비교에서는 남자가 여자보다 조금 더 유의했지만 그 외 나머지 비교에서는 여자가 남자보다 훨씬 더 유의했다. 남자의 경우는 $G_2 vs. G_3$ 의 비교

표 3.2. 제한한 통계량 P 와 Sun (2006)의 통계량 S 의 값과 $B = 50$ 번의 반복 대체에 의해 얻어진 유의확률 값 (G_1 : 20-29세, G_2 : 30-39세, G_3 : 40-49세, G_4 : 50세 이상)

비교연령그룹	남자		여자	
	P	S	P	S
G_1 vs. G_2	9.44 (0.002)	5.76 (0.016)	7.34 (0.007)	3.12 (0.077)
G_1 vs. G_3	75.04 (< 0.001)	66.96 (< 0.001)	93.46 (< 0.001)	74.85 (< 0.001)
G_1 vs. G_4	49.92 (< 0.001)	31.57 (< 0.001)	157.86 (< 0.001)	119.39 (< 0.001)
G_2 vs. G_3	130.80 (< 0.001)	134.69 (< 0.001)	142.94 (< 0.001)	145.77 (< 0.001)
G_2 vs. G_4	51.59 (< 0.001)	41.42 (0.001)	213.13 (< 0.001)	196.45 (< 0.001)
G_3 vs. G_4	5.54 (0.019)	9.098 (0.003)	19.92 (< 0.001)	14.55 (< 0.001)
모든 G_i , $i = 1, \dots, 4$	181.14 (< 0.001)	170.48 (< 0.001)	327.45 (< 0.001)	291.79 (< 0.001)

가 가장 유의했지만 여자의 경우는 G_2 vs. G_4 의 비교가 가장 유의했다. 한편, 4개 연령그룹의 생존함수를 동시에 비교하였다. 남자의 경우는 두 검정통계량의 값이 각각 $P = 181.14$ (유의확률 값 < 0.001), $S = 170.48$ (유의확률 값 < 0.001)로 4개 연령그룹 간 생존율의 차이가 매우 큼을 알 수 있었다. 여자의 경우도 4개 연령그룹 간 차이가 있는 것으로 나타났는데($P = 327.45$, $S = 291.79$) 남자의 경우보다 그 차이가 훨씬 더 유의했다.

4. 모의실험

2.2절과 2.3절에서 소개한 통계량 S 와 P 의 소표본 성질을 비교하기 위해 모의실험을 수행했다. 모의실험 설계는 Pan (2001)과 유사하게 했다. 두 집단($p = 2$)의 생존함수를 비교하기 위해 각 집단으로부터 50개의 자료를 추출했다($n = 100$). 감염시점은 로그정규분포 $LN(3.8, 0.3)$ 로부터 생성했다. 검진연구의 패턴과 유사하도록 하기 위해 최초 검진시점은 균등분포 $U(0, 5)$ 에서 생성했고, 그 이후는 한 달 간격으로 검진이 이루어지도록 계획했다. 계획된 시점에서 실제로 검진이 이루어지는지 여부는 베르누이 분포 $Bernoulli(\pi)$ 에 따른다. 단, $\pi \in (0, 1)$ 는 성공률이다. 따라서 감염시점은 연속된 두 방문시점에 의해 구간중도절단 된다. 모의실험에서 고려한 두 베르누이분포는 $Bernoulli(0.5)$ 와 $Bernoulli(0.3)$ 이다. 구간중도절단 되는 비율은 전자보다 후자가 크며, 평균구간중도절단 길이는 각각 2.9(달), 4.2(달)이다. 한편, 각 그룹의 잠복시간은 형상모수(shape parameter)가 동일한 와이블분포 $W(\alpha, \beta)$ 로부터 생성했다. 단, $\alpha(> 0)$ 는 형상모수이고 $\beta(> 0)$ 는 척도모수(scale parameter)이다. 모의실험에서 고려한 세 가지 다른 조합은 (i) $W(2.5, 60)$ vs. $W(2.5, 60)$, (ii) $W(2.5, 60)$ vs. $W(2.5, 70)$, (iii) $W(2.5, 60)$ vs. $W(2.5, 79)$ 이다. 각 조합별 위험률의 비는 (i) $HR = 1.00$, (ii) $HR = 1.47$, (iii) $HR = 1.99$ 이다. 질병의 발현시점은 랜덤하게 우중도절단 시켰으며, 그 비율이 $RF = 10\%$, $RF = 30\%$ 가 되게 했다. 1,000번 반복하였으며, 분산 추정을 위해 $B = 20$ 번 반복 대체했다.

표 4.1의 결과를 살펴보면 먼저 두 검정통계량은 모두 구간중도절단 되는 비율이나 우중도절단 되는 비율(RF)에 관계 없이 명목 유의수준 0.05를 잘 조절하는 것으로 나타났다($HR = 1$ 에 해당). 검정력 측

표 4.1. 1,000번 반복과 $B = 20$ 번 반복 대체에 의한 검정통계량 S 와 P 의 실험적 유의수준과 검정력 비교

RF(%)	HR	Bernoulli(0.5)		Bernoulli(0.3)	
		S	P	S	P
10	1.00	0.047	0.051	0.045	0.041
	1.47	0.411	0.423	0.428	0.444
	1.99	0.888	0.897	0.878	0.885
30	1.00	0.055	0.057	0.056	0.052
	1.47	0.340	0.349	0.352	0.364
	1.99	0.767	0.786	0.799	0.822

면에서 두 검정법은 모두 우중도절단 되는 비율이 증가할수록 감소했으나 구간중도절단 되는 비율에 따른 일관된 경향은 없었다(HR = 1.47, HR = 1.99에 해당). 두 검정법의 비교에서는 구간중도절단 되는 비율이나 우중도절단 되는 비율에 관계 없이 P 의 검정력이 S 의 검정력보다 높게 나타났다.

5. 결론

이중구간중도절단된 자료의 생존함수를 비교하기 위한 두 검정법을 소개하고, 한국인 암 예방연구 자료에 적용하여 당뇨병 잠복시간의 분포를 성별과 연령에 따라 비교하였다. Kim 등 (2006)의 통계량을 이중구간중도절단된 자료로 확장한 통계량 P 는 위험집합의 크기에만 의존하는 가중값을 사용하기 때문에 Sun (2006)처럼 X 와 T 에 대한 생존함수의 추정을 필요로 하지 않는다. 따라서 제안한 검정법은 KCPS 자료와 같이 대용량 자료의 분석에서 계산 시간을 대폭 줄일 수 있고, 또한 Sun (2006)이 지적한 것처럼 n 이 매우 크거나 연속형 생존자료에 대해 De Gruttoda과 Lagakos (1989)의 자기일치 추정량은 수렴하지 않을 수도 있기 때문에 Sun (2006)의 검정법은 이산형 생존자료의 분석에 적합한 반면, 제안한 검정법은 이산형 자료뿐만 아니라 연속형 자료에도 적용 가능한 장점이 있다. 로그-순위형 통계량의 공분산행렬을 추정하기 위해 2.2절과 2.3절에서 도입한 다중대체 방법의 대안으로 붓스트랩 방법과 Markov Chain Monte Carlo(MCMC) 방법을 고려 할 수 있다 (Zhao와 Sun, 2004). 전자의 방법은 단순하여 적용하기가 쉽지만 우중도절단된 자료에 대해 로그-순위 통계량의 추정 공분산행렬을 얻지 못하는 단점을 갖고 있다. 후자의 방법도 붓스트랩 방법과 마찬가지로 우중도절단된 자료의 경우 로그-순위 통계량의 추정 공분산행렬로 표현되지 않으며, 다중대체 방법보다 계산 시간이 많이 요구되는 단점을 갖고 있다. 반면에 다중대체 방법은 우중도절단된 자료에 대해 \hat{V}^* 가 로그-순위 통계량의 추정 공분산행렬로 축소되는 장점을 갖고 있다. 사용 편리성 측면에서는 다중대체 방법이 붓스트랩 방법보다는 복잡하고 MCMC 방법보다는 간편하다고 말할 수 있다.

KCPS 자료를 분석할 때 여자의 경우는 모든 개체를 분석에 포함했지만, 남자의 경우는 연령과 우중도절단 여부에 따라 나는 8개 층에서 추출률을 0.25로 하여 얻은 자료만을 분석에 포함했다. 3절에서 언급한 5가지 조건 중 어느 것에도 해당되지 않는 남성들을 분석에 모두 포함시키지 않은 이유는 남녀 성비를 고려한 측면도 있지만 De Gruttoda과 Lagakos (1989)의 생존함수 추정이 가능해지도록 하기 위함이었다. 성별에 따라 당뇨병의 잠복시간이 매우 유의하게 달랐으며 여자의 잠복시간이 남자보다 긴 것으로 나타났다. 4개 연령그룹 간 당뇨병의 잠복시간도 성별에 관계 없이 매우 유의하게 달랐으며 여자의 경우가 남자의 경우보다 그 차이가 훨씬 더 유의했다. 두 검정통계량 S 와 P 를 서로 비교하면, 성별에 따른 잠복시간 분포의 차이가 S 보다 P 에서 더 유의했고, 성별 내에서 4개 연령그룹 간 생존률의 차이도 S 보다 P 에서 더 유의했다. 한편, 모의실험 결과를 보면 두 검정법 모두 명목 유의수준을 잘 조절하였으며 제안한 검정법의 검정력이 Sun (2006)의 검정력보다 높게 나타났다.

2.2절과 2.3절에서 다룬 검정법들은 모두 무절단(no truncation)을 가정하고서 얻어진 것이다. 그런데 전향적 연구(prospective study)에서는 어떤 개체의 시작사건이 연구기간 전에 발생하면 그 개체는 좌절단(left-truncated)되고, 후향적 연구(retrospective study)에서는 사망사건이 연구기간 이후에 발생하면 우절단(right-truncated)되기 때문에 절단된 자료를 무시하고 분석하면 샘플링 편의(length bias sampling)가 발생할 수 있다. 이와 같은 문제를 해결하기 위해 향후 연구에서는 Kim 등 (2006)의 검정법을 구간중도절단 뿐만 아니라 절단을 포함하는 자료로 확장하고자 한다.

참고문헌

- 김진화, 한미아, 박철진, 박일구, 신지혜, 김상용, 류소연, 배학연 (2008). 나주지역 중년인구에서 공복혈당을 통한 당뇨병 진단의 적절성 평가, <대한당뇨병학회>, **32**, 328-337.
- 박이병 (2006). Prediabetes-risk of type 2 diabetes and CVD, <제19차 대한당뇨병학회 춘계학술대회>.
- 통계청 (2007). <2006년 사망원인 통계연보>, 통계청.
- De Gruttola, V. G. and Lagakos, S. W. (1989). Analysis of doubly-censored survival data, with application to AIDS, *Biometrics*, **45**, 1-11.
- Fay, M. P. (1996). Rank invariant tests for interval censored data under the grouped continuous model, *Biometrics*, **52**, 811-822.
- Fonseca, V. A. (2007). Identification and treatment of prediabetes to prevent progression to type 2 diabetes, *Clinical Cornerstone*, **8**, 10-18.
- Gómez, G. and Lagakos, S. W. (1994). Estimation of the infection time and latency distribution of AIDS with doubly censored data, *Biometrics*, **50**, 204-212.
- Jee, S. H., Ohrr, H., Sull, J. W., Yun, J. E., Ji, M. and Samet, J. M. (2005). Fasting serum glucose level and cancer risk in Korean men and women, *Journal of the American Medical Association*, **293**, 194-202.
- Kim, J., Kang, D. R. and Nam, C. M. (2006). Logrank-type tests for comparing survival curves with interval-censored data, *Computational Statistics and Data Analysis*, **50**, 3165-3178.
- Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, Springer, New York.
- Korea Centers for Disease Control and Prevention (2007). <국민건강영양조사 제3기 조사결과 심층분석 연구보고서: 김진부문>, 질병관리본부.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, John Wiley & Sons, New York.
- Pan, W. (2000). A two-sample test with interval censored data via multiple imputation, *Statistics in Medicine*, **19**, 1-11.
- Pan, W. (2001). A multiple imputation approach to regression analysis for doubly censored data with application to AIDS studies, *Biometrics*, **57**, 1245-1250.
- Self, S. G. and Grossman, E. A. (1986). Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers, *Biometrics*, **42**, 521-530.
- Sun, J. (1996). A nonparametric test for interval-censored failure time data with application to AIDS studies, *Statistics in Medicine*, **15**, 1387-1395.
- Sun, J. (1997). Self-consistency estimation of distributions based on truncated and doubly censored data with applications to AIDS cohort studies, *Lifetime Data Analysis*, **3**, 305-313.
- Sun, J. (2001). Nonparametric test for doubly interval-censored failure time data, *Lifetime Data Analysis*, **7**, 363-375.
- Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*, Springer, New York.
- Zhao, Q. and Sun, J. (2004). Generalized log-rank test for mixed interval-censored failure time data, *Statistics in Medicine*, **23**, 1621-1629.

Comparing Survival Functions with Doubly Interval-Censored Data: An Application to Diabetes Surveyed by Korean Cancer Prevention Study

Sun Ha Jee¹ · Chung Mo Nam² · Jinheum Kim³

¹Institute for Health Promotion, Graduate School of Public Health Yonsei University;

²Department of Preventive Medicine, Yonsei University College of Medicine;

³Department of Applied Statistics, University of Suwon

(Received January 2009; accepted March 2009)

Abstract

Two tests were introduced for comparing several survival functions with doubly interval-censored data and illustrated with data surveyed by Korean Cancer Prevention Study (Jee *et al.*, 2005). The test which extended Kim *et al.* (2006)'s test to the doubly interval-censored data has an advantage over Sun (2006)'s test in terms of saving computation time because the proposed test only depends on the size of risk set, and also the proposed test is applicable to continuous failure time data as well as discrete failure time data unlike Sun's test. Comparing male with female groups on the incubation time of diabetes was highly different and the survival of female group was longer than that of male one. Regardless of gender, the difference in survival functions of four age groups was highly significant with p -value of less than 0.001. This trend was more remarkable for female group than for male one. Simulation results showed that the significance level of both tests was well controlled and the proposed test was better than Sun's test in terms of power.

Keywords: Doubly interval-censored, survival function, log-rank test, multiple imputation, incubation time.

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD)(KRF-2007-314-C00042).

³Corresponding author: Associate Professor, Department of Applied Statistics, University of Suwon, San 2-2 Wau-Ri, Bongdam-Eup, Whasung 445-743, Korea. E-mail: jinhkim@suwon.ac.kr