Basado en la literatura del corpus, trabajar con muestras pequeñas ($N < 400$) y alta censura (más del 50%) presenta desafíos metodológicos específicos que afectan tanto la elección del modelo como la interpretación de los resultados.
Aquí detallo las limitaciones estadísticas y las métricas recomendadas para este escenario:

1. Limitaciones Estadísticas con $N < 400$ y Alta Censura

A. Reducción del Tamaño de Muestra Efectivo (Eventos por Variable) La limitación más crítica no es el tamaño total de la muestra (N), sino el número de **eventos observados**. Con un $N=400$ y una censura $>50\%$, tienes menos de 200 eventos efectivos.

- **Regla empírica:** En estudios clínicos y epidemiológicos (aplicables aquí), se recomienda mantener una relación adecuada entre el número de eventos y el número de predictores (covariables) para evitar el sobreajuste (*overfitting*). Spoto et al. sugieren, por ejemplo, introducir aproximadamente una covariable por cada 22 eventos para mantener una potencia estadística adecuada 1.
- **Consecuencia:** Si intentas usar muchos predictores (ej. múltiples competencias blandas, datos demográficos y académicos) con menos de 200 eventos, los modelos complejos (como Deep Learning o incluso un Cox multivariante denso) tendrán coeficientes inestables y baja capacidad de generalización.

B. Inestabilidad de Modelos de Machine Learning (ML) Aunque modelos como *Random Survival Forests* (RSF) son potentes, la literatura advierte sobre su uso en muestras pequeñas.

- **RSF:** Garcia-Lopez et al. señalan explícitamente que, aunque el RSF tiene ventajas sobre los métodos clásicos, presenta una **precisión predictiva reducida en muestras pequeñas** 2.
- **Redes Neuronales:** Kvamme y Borgan demuestran que las redes neuronales para supervivencia requieren una cuidadosa discretización del tiempo. Con muestras pequeñas, el uso de rejillas de tiempo finas aumenta el número de parámetros, haciendo que la red sea propensa al sobreajuste (*overfitting*) 3, 4.
- **DeepHit:** Rossi et al. encontraron que modelos de aprendizaje profundo como *DeepHit* rinden mal en tamaños de muestra más bajos y muestran una calibración deficiente en comparación con modelos más simples 5, 6.

C. El Fenómeno de la "Fracción de Cura" (Cure Fraction) En estudios de mercado laboral con alta censura, existe el riesgo de confundir "censura" (no encontró trabajo *todavía*) con "inmunidad" o "cura" (nunca buscará trabajo o nunca lo encontrará, por ejemplo, por decidir ser ama de casa o salir de la fuerza laboral).

- **Evidencia:** Abdullahi et al. argumentan que cuando la censura es alta (en su caso 33.8% de desempleo), los modelos tradicionales (Cox/AFT) pueden no ser apropiados porque asumen que todos eventualmente experimentarán el evento. En estos casos, los **Modelos de Cura Mixta (Mixture Cure Models)** son superiores para distinguir entre quienes tardan en encontrar empleo y quienes están estructuralmente desempleados 7, 8.

2. Métricas Aceptables e 'Informativas'

Dado el contexto de alta censura, no basta con medir la discriminación (el orden de los eventos); es crucial medir la calibración (la precisión de la probabilidad).

A. Concordance Index (C-index) de Antolini (No el de Harrell)

- **Limitación de Harrell:** El C-index estándar de Harrell asume que el ranking de riesgo entre individuos se mantiene constante en el tiempo. Sin embargo, en modelos de ML o datos complejos, las curvas de supervivencia pueden cruzarse (violación de riesgos proporcionales).
- **Recomendación:** Rossi et al. recomiendan encarecidamente usar el **C-index de Antolini**, ya que generaliza la métrica para casos donde el supuesto de riesgos proporcionales (PH) no se cumple, lo cual es común en datos de empleo complejos 9, 10.

B. Integrated Brier Score (IBS) para Calibración Un modelo puede ordenar bien a los candidatos (alto C-index) pero dar probabilidades de empleo totalmente erróneas.

- **Evidencia:** Suresh et al. y Rossi et al. destacan el **Brier Score (BS)** y el **Integrated Brier Score (IBS)** como métricas esenciales. El IBS evalúa simultáneamente la discriminación y la calibración a lo largo del tiempo, ponderando por la inversa de la probabilidad de censura (IPCW) para corregir el sesgo introducido por los sujetos censurados 11, 12. Un IBS bajo indica un modelo robusto y bien calibrado.

C. AUC Dependiente del Tiempo (Time-dependent AUC) Para estudios psicosociales o de empleo, a menudo interesa una predicción en un horizonte específico (ej. "¿Probabilidad de empleo a los 6 meses?").

- **Uso:** Suresh et al. sugieren el uso del AUC dependiente del tiempo (Cumulative/Dynamic AUC) para evaluar la capacidad del modelo de distinguir entre quienes encuentran empleo antes de un tiempo t y quienes no 13. Esto es más informativo para la toma de decisiones políticas o educativas que un índice global.

Resumen de Recomendación

Para tu tesis con $N < 400$ y alta censura:

1. **Prioriza la parsimonia:** Evita modelos de Deep Learning complejos (DeepHit); opta por modelos paramétricos (AFT Weibull/Log-Normal) o Cox penalizado (Lasso/Elastic Net) si tienes muchas variables 14.
2. **Valida con IBS:** No reportes solo el C-index. Reporta el **Integrated Brier Score (IBS)** para demostrar que tu modelo no solo ordena bien a los graduados, sino que sus probabilidades de tiempo son realistas 15.
3. **Considera la fracción de cura:** Si sospechas que muchos graduados no buscarán empleo activamente, justifica si un modelo de cura (Mixture Cure Model) ajustaría mejor que un Cox estándar 7.