2026 Little Hearts Explorer | 아이마음 탐사대

발달지연 아동을 위한 효과적 조기개입 솔루션 발굴

SPACE 2 임상연구 설계 기준 보고서

과학적으로 검증된 근거 생성을 위한 도메인별 연구 설계 가이드라인

SPACE 2 Clinical Research Design Standards for Identifying Effective Early Intervention Solutions

---

연구 기관	연세대학교 인구와인재연구원
연구 총괄	김현철 교수
연구 담당	나성실 연구교수
문서 버전	Ver. 1.1

본 보고서는 아이마음 탐사대 SPACE 2 참여 기관이 발달지연 및 발달장애 아동을 위한 조기개입 솔루션의 효과를 과학적으로 검증하는 데 필요한 임상연구 설계 기준을 제시하기 위해 작성된 가이드라인 문서입니다. 본 문서에 제시된 설계 기준은 참여 기관의 연구 역량 및 솔루션 특성에 따라 자문단과의 협의를 통해 조정될 수 있습니다.

참고 기준 및 지침

본 보고서는 아래의 문헌에 근거하여 작성됨.

- 한국보건의료연구원(NECA). 신의료기술평가 체계적 문헌고찰 지침. 2011.
- 한국보건의료연구원(NECA). 알기 쉬운 신의료기술평가 제도 A to Z. 2025.
- 한국보건의료연구원(NECA). 2024 년 신의료기술평가 사례 분석 보고서. 2024.
- 건강보험심사평가원(HIRA). 근거기반 의사결정을 위한 근거문헌활용지침 제 7 판. 2024.
- Higgins et al. Revised Cochrane Risk-of-Bias Tool for Randomized Trials (RoB 2). 2019.
- Hopewell et al. CONSORT 2025 Statement. JAMA, 2025.
- SPIRIT 2013 Statement.
- GRADE (Grading of Recommendations Assessment, Development and Evaluation) Framework.

서 문

아이마음탐사대 프로젝트

아이마음탐사대 (Little Hearts Explorer) 프로젝트 SPACE 2 는 "발달지연 아동을 위한 효과적 조기개입 솔루션 발굴"을 목표로, SPACE 1 에 진출한 12 개 우수 연구팀 중 최종 선발된 6 개 팀이 참여하게 됩니다. 각 팀은 부모 매개 중재, 디지털 치료기기, 언어·말소리 치료, 인지재활, 학교준비 프로그램, 진단 도구 등 다양한 영역에서 발달장애 아동을 위한 혁신적 솔루션을 개발하고 있습니다. 본 프로젝트의 궁극적 목표는 이들 솔루션이 과학적으로 검증되어 실제 임상 현장에 도입됨으로써, 모든 발달장애 아동이 가정의 경제적 여건이나 거주 지역과 무관하게 효과적인 치료를 받을 수 있는 사회를 만드는 것입니다. 그러나 아무리 좋은 의도로 개발된 치료법이라도 과학적 근거 없이는 사회적 신뢰를 얻기 어렵고, 실제로 도움이 필요한 아이들과 가족에게 널리 확산되지 못합니다. 본 가이드라인은 SPACE 2 를 준비하는 모든 연구팀, 그리고 최종 선발되어 참여하게 될 6 개 팀이 과학적으로 엄격하게 검증된 근거를 생성하여 개발한 솔루션이 실제로 아동과 가족에게 닿을 수 있도록 돕기 위해 개발되었습니다.

이 가이드라인이 제공하는 것

본 가이드라인은 SPACE 2 를 참여하는 모든 연구팀이 반복되는 실패를 피하고 처음부터 올바른 방향으로 연구를 설계할 수 있도록 체계적이고 실용적인 지침을 제공합니다. 연구 설계 유형부터 대상자 선정, 표본 크기 산출, 중재 프로토콜, 결과 변수, 평가 시점, 눈가림, 데이터 관리, 통계 분석, 보건경제성, 윤리, 정책 연계까지 임상연구의 전 과정을 12 개 Domain 으로 체계화했습니다. 각 Domain 은 단순한 이론 설명이 아니라 "연구팀이 해야 할 것"을 중심으로 구체적이고 실행 가능한 지침을 제시하며, 누구나 이해할 수 있는 쉬운 언어로 작성하되 과학적 엄격성은 절대 타협하지 않았습니다. 또한 부록 A(진단검사 정확도 연구), 부록 B(중재 유형별 예시와 자주 발생하는 오류), 부록 C(검증된 표준화 검사도구)를 통해 실무에서 바로 적용할 수 있는 구체적 도구를 제공합니다.

우리가 궁극적으로 바라는 것

본 가이드라인의 최종 목표는 발달장애 아동과 가족의 삶을 실질적으로 변화시키는 것입니다. 연구팀들이 생성하는 과학적 근거는 단순히 학술적 성과가 아니라, 모든 아동이 태어난 지역이나 가정의 경제적 형편과 무관하게 효과적인 치료를 받을 수 있는 사회를 만들기 위한 필수적 과정입니다. 디지털 치료기기와 부모 매개 중재를 통해 대도시 대형병원에 가지 않아도, 농어촌이나 의료 소외 지역에 살아도, 집에서 부모와 함께 효과적인 치료를 받을 수 있는 새로운 패러다임을 만들고자 합니다. 더 나아가 본 프로젝트를 통해 확립된 연구

방법론이 향후 국내에서 개발되는 수많은 혁신적 치료법들의 표준이 되어, 진정으로 효과 있는 치료들이 과학적으로 검증되고 사회 전체로 확산되는 선순환 구조를 만들고자 합니다. 그러나 무엇보다 우리가 진정으로 바라는 것은 언어 발달이 늦은 아이가 처음으로 "엄마"라고 말하는 순간, 학교 적응이 어려웠던 아이가 친구들과 함께 웃으며 노는 모습, 부모가 "우리 아이도 할 수 있구나"라는 희망을 갖는 것입니다.

연구팀 여러분은 탐험가입니다

본 가이드라인을 손에 든 연구팀 여러분은 단순히 연구자가 아니라, 수만 명의 발달장애 아동과 가족에게 희망을 전하는 탐험가(Explorer)입니다. 여러분이 설계하는 연구 하나하나, 수집하는 데이터 하나하나, 분석하는 결과 하나하나가 실제 아동의 삶을 변화시킬 수 있는 근거가 되며, 방법론적으로 완벽하지 않은 연구는 단순히 학술지에 실리지 못하는 것을 넘어 도움을 기다리는 아이들에게 효과적인 치료가 전달되지 못하는 결과를 초래합니다. 본 가이드라인은 여러분이 그 무거운 책임을 홀로 짊어지지 않도록 돕기 위해 만들어졌으며, 각 Domain 의 지침을 따라가다 보면 여러분은 어느새 국제 수준의 임상연구를 설계하고 수행하고 있을 것입니다. 어려운 결정을 내려야 할 때, 복잡한 방법론적 문제에 직면했을 때, 이 가이드라인으로 돌아오십시오. 여러분은 혼자가 아니며, 모두가 같은 목표를 향해 함께 나아가고 있습니다.

과학적 근거 생성은 결코 쉬운 일이 아니지만, 이 길이 발달장애 아동과 가족에게 진정으로 도움이 되는 유일한 길임을 우리는 압니다. 본 가이드라인이 여러분의 여정에 든든한 나침반이 되기를, 그리고 SPACE 2 프로젝트를 통해 생성된 근거들이 대한민국의 발달장애 아동 치료의 패러다임을 바꾸고 나아가 전 세계 발달장애 아동에게 희망을 전하는 이정표가 되기를 간절히 소망합니다.

연세대학교 인구와인재연구원

원장 김현철

목 차

서 문 iii

약어 정리 1

Domain 1. 연구 설계 유형 (Study Design) 3

Domain 2. 연구대상자 선정 및 무작위배정 (Participant Selection & Randomization) 6

Domain 3. 표본 크기 산출 (Sample Size Calculation) 10

Domain 4. 중재 프로토콜 표준화 (Intervention Protocol Standardization) 14

Domain 5. 결과 변수 선정 (Outcome Measures) 18

Domain 6. 평가 시점 및 평가 운영 (Assessment Timing & Operations) 22

Domain 7. 평가자 관리 및 눈가림 (Assessor Management & Blinding) 25

Domain 8. 데이터 입력 및 품질 관리 (Data Collection & Quality Management) 29

Domain 9. 통계 분석 계획 (Statistical Analysis Plan) 32

Domain 10. 보건경제성 분석 (Health Economic Analysis) 35

Domain 11. 윤리 및 규제 준수 (Ethics & Regulatory Compliance) 38

Domain 12. 일반화 가능성 및 정책 연계 전략 (Generalizability & Policy Translation) 41

전체 도메인 통합 요약표 45

부록 A. 진단검사 정확도 연구 가이드라인 46

부록 B. 중재 유형별 예시 모음 72

1. 부모 매개 중재 (Parent-Mediated Intervention) 75

2. 디지털 중재 (Digital Intervention) 84

3. 언어, 말소리 중재 90

4. 인지, 실행기능 재활 93

5. 학교준비 프로그램 95

부록 C. 평가영역별 표준화 검사도구 96

약어 정리

약어	Full Term	뜻
ADHD	Attention Deficit Hyperactivity Disorder	주의력결핍 과잉행동장애
ASD	Autism Spectrum Disorder	자폐스펙트럼장애
AUC	Area Under the Curve	ROC 곡선 아래 면적
BRIEF / K-BRIEF	Behavior Rating Inventory of Executive Function	한국판 실행기능 행동평가 척도
CBCL	Child Behavior Checklist	아동·청소년 행동평가척도
CONSORT	Consolidated Standards of Reporting Trials	무작위 임상시험 보고 통합 기준
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition	정신질환 진단 및 통계 편람 제 5 판
DSMB	Data Safety Monitoring Board	자료안전성 모니터링 위원회
DTA	Diagnostic Test Accuracy	진단검사 정확도
GRADE	Grading of Recommendations Assessment, Development and Evaluation	권고안 평가·개발·평가 등급화 체계
HIRA	Health Insurance Review & Assessment Service	건강보험심사평가원
ICC	Intraclass Correlation Coefficient	급내상관계수
ICD-11	International Classification of Diseases, 11th Revision	국제질병분류 제 11 차 개정판
IRB	Institutional Review Board	기관생명윤리위원회
ITT	Intention-to-Treat	치료 의도 분석
K-WISC-V	Korean Wechsler Intelligence Scale for Children, 5th	한국 웩슬러 아동 지능검사 5 판
MLU	Mean Length of Utterance	평균 발화 길이

NECA	National Evidence-based healthcare Collaborating Agency	한국보건의료연구원
NPV	Negative Predictive Value	음성 예측도
PKBS-2	Preschool and Kindergarten Behavior Scales, 2nd	유치원 행동 척도 2 판
PPV	Positive Predictive Value	양성 예측도
PSI-SF	Parenting Stress Index-Short Form	부모 양육 스트레스 척도 단축형
RCT	Randomized Controlled Trial	무작위 대조 임상시험
REVT	Receptive & Expressive Vocabulary Test	수용·표현 어휘력 검사
RoB	Risk of Bias	비뚤림 위험
ROC	Receiver Operating Characteristic curve	수신자 조작 특성 곡선
SSIS	Social Skills Improvement System	사회성 기술 향상 체계
SRS-2	Social Responsiveness Scale, 2nd	사회적 반응성 척도 2 판
TAU	Treatment As Usual	일반적 치료
VABS-II	Vineland Adaptive Behavior Scales, 2nd	바인랜드 적응행동 척도 2 판

Domain 1. 연구 설계 유형 (Study Design)

연구 설계란 무엇인가요?

연구 설계는 우리가 개발한 치료법이 정말 효과가 있는지 확인하는 방법을 미리 계획하는 것입니다.

왜 비교 그룹이 필요한가요?

발달이 늦은 아이들은 특별한 치료를 받지 않아도 시간이 지나면서 저절로 좋아지기도 합니다. 따라서 "치료 전"과 "치료 후"만 비교하면, 아이가 좋아진 이유가 우리 치료법 덕분인지 아니면 그냥 나이를 먹으면서 자연스럽게 발달한 것인지 구분할 수 없습니다.

이 문제를 해결하려면 반드시 비교 그룹(대조군)이 필요합니다. 무작위 대조 임상시험(RCT)은 참여 아동을 제비뽑기처럼 공정하게 두 그룹으로 나누어, 한 그룹은 우리 치료를 받고 다른 그룹은 일반적인 치료를 받게 한 뒤 결과를 비교하는 방식입니다.

연구팀이 준비해야 할 것

연구를 시작하기 전에 다음 사항을 문서로 작성하여 확정해야 합니다:

- RCT 기반의 연구 구조
- 한 곳에서 할지 여러 기관에서 함께 할지
- 치료는 몇 주 동안 진행할지
- 치료 후 얼마나 오래 추적 관찰할지

근거 수준별 연구 설계 방법 1

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• RCT 기반 연구 구조 채택 • 단일기관 또는 다기관 여부 사전 결정 및 명시 • 연구 기간(중재 기간, 추적관찰 기간) 및 전체 구조 사전 정의	Confounding bias
상 (권장)	• Pragmatic RCT 설계 적용 (예: 실제 임상 환경 반영, 일반화 가능성 강화) • 다기관 설계: 최소 3 개 이상 기관 참여 • Active comparator 설계 도입 (예: 기존 표준치료와의 직접 비교)	
최상 (이상)	• Cluster RCT 또는 Stepped-wedge 설계 검토 (예: 기관 단위 무작위배정) • Adaptive design 적용 (예: 중간 분석 결과에 따른 표본 수 조정 계획 사전 정의)	

최소 충족 기준 요약: RCT 기반 설계 채택 + (해당 시) 기관 수 명시 + 연구 기간 및 전체 구조 사전 정의

자가 점검 체크리스트 1

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	연구 설계를 RCT 로 확정되어 있다	☐	☐	☐
🔴	단일기관 또는 다기관 여부를 결정하고 문서에 명시하였다	☐	☐	☐
🔴	중재 기간과 추적관찰 기간을 포함한 전체 연구 기간을 사전에 정의하였다	☐	☐	☐
🟡	실제 임상 환경을 반영한 Pragmatic RCT 설계를 적용하였다	☐	☐	☐
🟡	3 개 이상의 기관이 참여하는 다기관 설계를 구성하였다	☐	☐	☐
🟡	기존 표준치료와 직접 비교하는 Active comparator 설계를 도입하였다	☐	☐	☐
🟢	Cluster RCT 또는 Stepped-wedge 설계 적용 가능성을 검토하였다	☐	☐	☐
🟢	중간 분석 결과에 따른 표본 수 조정이 가능한 Adaptive design 계획을 수립하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 3 개

Domain 2. 연구대상자 선정 및 무작위배정 (Participant Selection & Randomization)

누가 연구에 참여하고, 어떻게 그룹을 나눌까요?

연구대상자 선정 및 무작위배정은 두 가지를 정하는 과정입니다:

1. 어떤 아이들이 연구에 참여할 수 있는지 (포함/제외 기준)
2. 참여하는 아이들을 치료 그룹과 비교 그룹에 어떻게 공정하게 나눌지

왜 중요한가요?

만약 선정 기준이 애매하거나 연구자가 마음대로 그룹을 정하면, 두 그룹이 처음부터 다르게 구성될 수 있습니다. 예를 들어 증상이 가벼운 아이들은 치료 그룹에, 증상이 심한 아이들은 비교 그룹에 배정된다면, 나중에 결과가 좋아 보이는 이유가 치료 효과가 아니라 애초에 두 그룹이 달랐기 때문일 수 있습니다. 이런 문제는 연구가 끝난 후에는 통계로도 바로잡기 어렵습니다.

연구팀이 해야 할 것

1. **명확한 기준 설정**: DSM-5 또는 ICD-11 같은 공인된 진단 기준을 사용하여 누가 참여할 수 있고 없는지 구체적으로 정합니다.
2. **공정한 그룹 나누기**: 컴퓨터가 무작위로 그룹을 배정하도록 하고, 연구자도 다음 아동이 어느 그룹으로 갈지 미리 알 수 없도록 배정 결과를 숨겨둡니다(배정 은폐).

근거 수준별 연구 설계 방법 2

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 명확한 포함/제외 기준 설정 (예: 진단 기준 단일화: DSM-5 또는 ICD-11 기반) • 적절한 비교군 구조 확보 (예: Waitlist control 또는 TAU 중 하나 이상) • Random sequence generation 적용 (예: 컴퓨터 기반 난수 생성) • Allocation concealment 적용	Selection bias, Randomization bias
상 (권장)	• 대상자 모집 경로 및 선정 절차 표준화 (예: 다기관 공통 스크리닝 프로토콜) • Baseline comparability 확보 (예: 배정 전 기저 특성 동질성 검증) • Stratified/random block 설계 적용 (층화변수: 연령, 중증도, 기관 등)	
최상 (이상)	• 모집 및 배정 과정의 사전 정의와 재현 가능성 확보 • 무작위배정 전 과정 문서화 (CONSORT 흐름도 포함)	

최소 충족 기준 요약: 포함/제외 기준 명문화 + 비교군 설정 + Random sequence generation + Allocation concealment

자가 점검 체크리스트 2

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	포함 기준과 제외 기준을 DSM-5 또는 ICD-11 기반으로 명확히 정의하였다	☐	☐	☐
🔴	Waitlist control 또는 TAU 중 하나 이상의 비교군을 확보하였다	☐	☐	☐
🔴	컴퓨터 기반 난수 생성 방식으로 무작위배정 순서를 사전에 생성하였다	☐	☐	☐
🔴	배정 결과를 아동 등록 시까지 연구자가 알 수 없도록 배정 은폐 절차가 갖춰져 있다	☐	☐	☐
🟡	다기관 공통 스크리닝 프로토콜을 수립하여 모집 경로와 선정 절차를 표준화하였다	☐	☐	☐
🟡	배정 전 두 군의 기저 특성 동질성을 확인하는 절차를 계획하였다	☐	☐	☐
🟡	연령, 중증도, 기관 등 층화변수를 적용한 Stratified/random block 설계를 적용하였다	☐	☐	☐
🟢	모집 및 배정 과정 전반을 사전 정의하고 재현 가능한 형태로 문서화하였다	☐	☐	☐
🟢	CONSORT 흐름도를 포함하여 무작위배정 전 과정을 문서화하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 4 개

자가 점검 체크리스트 2

(부모 매개 중재 연구용 추가 체크리스트)

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	양육자 포함/제외 기준을 별도로 정의하였다	☐	☐	☐
🔴	주 양육자가 프로그램 전 기간 참여 가능한지 확인하는 절차를 수립하였다	☐	☐	☐
🔴	양육자의 한국어 의사소통 능력을 포함/제외 기준에 명시하였다	☐	☐	☐
🔴	양육자의 심각한 정신질환, 약물/알코올 중독을 제외 기준에 포함하였다	☐	☐	☐
🟡	가정 환경의 최소 안전 기준(가정폭력 부재 등)을 제외 기준에 포함하였다	☐	☐	☐
🟡	다자녀 가구의 경우 프로그램 참여 가능성을 사전 평가하는 절차를 마련하였다	☐	☐	☐
🟢	양육자의 교육 수준, 사회경제적 지위를 층화변수로 고려하여 무작위배정하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 4 개

Domain 3. 표본 크기 산출 (Sample Size Calculation)

표본 크기란 무엇인가요?

표본 크기 산출은 치료 효과를 통계적으로 증명하기 위해 최소 몇 명의 아이들을 모집해야 하는지 미리 계산하는 과정입니다.

왜 미리 계산해야 하나요?

너무 적은 아이들로 연구하면, 치료 효과가 실제로 있어도 통계적으로 증명하지 못할 수 있습니다. 또한 연구 중에 일부 아이들은 이사나 개인 사정으로 그만둘 수 있는데(탈락), 이를 처음부터 고려하지 않으면 나중에 분석할 수 있는 아동 수가 부족해집니다.

예를 들어, 160 명이 필요한데 150 명만 모집했고 그중 30 명이 중도 탈락하면, 실제 분석 가능한 인원은 120 명에 불과하여 통계적 검정력이 부족해집니다.

연구팀이 해야 할 것

1. **효과 크기 추정**: 이전 비슷한 연구나 파일럿 데이터를 바탕으로 예상되는 치료 효과 크기를 확인합니다.

2. **표본 크기 계산**: 통계 프로그램(예: G*Power)을 사용하여 다음 기준으로 계산합니다.
    - 유의수준(α): 0.05
    - 검정력(Power): 80% 이상
3. **탈락률 반영**: 예상 탈락률(보통 15-20%)을 고려하여 초기 모집 목표를 더 높게 설정합니다.
    - 예: 필요 인원 160 명 → 탈락률 20% 고려 → 200 명 모집

근거 수준별 연구 설계 방법 3

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 표본 크기 산출 사전 수행 및 결과 보고 (유의수준 α=0.05, 검정력 80% 이상 기준) • 효과 크기 추정 근거 명시 (예: 선행 연구 또는 파일럿 데이터 기반) • 예상 탈락률 반영한 초기 모집 목표 수 설정 (통상 15~20% 추가 모집) • SPACE 2 최소 기준: 전체 160 명 이상 (case/control 각각 80 명 이상)	Attrition bias
상 (권장)	• 검정력 90% 기준 표본 크기 산출 (급여 심사용 보수적 기준 적용) • 탈락률 민감도 분석: 탈락률 10%, 20%, 30% 시나리오별 충분성 검토 • 층화변수별 하위군 분석 고려한 추가 표본 크기 검토	
최상 (이상)	• Adaptive sample size re-estimation 계획 수립 (예: 중간 분석 결과 기반 표본 수 조정) • 다기관 설계 시 기관 간 ICC 추정값 반영한 표본 크기 재산출 • 복수 1 차 결과 변수 설정 시 다중 비교 보정을 반영한 표본 크기 재산출	

최소 충족 기준 요약: Power analysis 수행 (α=0.05, Power≥80%) + 효과 크기 추정 근거 + 탈락률 반영 모집 수 설정 + SPACE 2 최소 대상자 수 (Case 군, control 군 각각 80 명)

자가 점검 체크리스트 3

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	유의수준 α=0.05, 검정력 80% 이상 기준으로 표본 크기 산출을 사전에 수행하였다	☐	☐	☐
🔴	선행 연구 또는 파일럿 데이터를 근거로 효과 크기 추정값이 문서에 명시되어 있다	☐	☐	☐
🔴	예상 탈락률(15~20%)을 반영하여 초기 모집 목표 수를 설정하였다	☐	☐	☐
🟡	급여 심사 보수적 기준 적용을 위해 검정력 90% 기준으로 표본 크기를 산출하였다	☐	☐	☐
🟡	탈락률 10%, 20%, 30% 시나리오별 충분성 검토를 수행하였다	☐	☐	☐
🟡	층화변수별 하위군 분석을 고려한 추가 표본 크기를 검토하였다	☐	☐	☐
🟢	중간 분석 결과에 따라 표본 수를 조정할 수 있는 Adaptive re-estimation 계획을 수립하였다	☐	☐	☐
🟢	다기관 설계 시 기관 간 ICC 추정값을 반영하여 표본 크기를 재산출하였다	☐	☐	☐
🟢	복수의 1 차 결과 변수 설정 시 다중 비교 보정을 반영한 표본 크기를 재산출하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 3 개

자가 점검 체크리스트 3

(부모 매개 중재 연구용 추가 체크리스트)

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	부모 매개 중재의 특성을 고려하여 탈락률을 가정하고 초기 모집 목표를 설정하였다	☐	☐	☐
🟡	과거 유사한 부모 매개 중재 연구의 탈락률 데이터를 근거로 제시하였다	☐	☐	☐
🟡	탈락 방지 전략(정기 연락, 교통비 지원, 유연한 일정)을 구체적으로 수립하였다	☐	☐	☐
🟢	양육자 특성별(맞벌이/외벌이, 다자녀/외동 등) 탈락률 차이를 민감도 분석으로 검토하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 1 개

Domain 4. 중재 프로토콜 표준화 (Intervention Protocol Standardization)

중재 프로토콜 표준화란 무엇인가요?

중재 프로토콜 표준화는 우리가 제공하는 치료의 모든 내용(무엇을, 어떻게, 몇 회, 얼마나)을 구체적으로 정하고 문서로 만드는 과정입니다.

왜 표준화가 필요한가요?

프로토콜이 명확하지 않으면 치료사마다, 기관마다, 회기마다 다른 치료를 제공하게 됩니다. 예를 들어, A 치료사는 주 3 회 놀이 중심으로 진행하고 B 치료사는 주 1 회 카드 학습 중심으로 한다면, 같은 "언어치료"라고 해도 실제로는 완전히 다른 치료입니다.

이렇게 되면 나중에 결과가 좋아도 그것이 우리 치료법 덕분인지, 아니면 단지 실력 좋은 치료사 덕분인지 알 수 없습니다. 또한 다른 기관에서 똑같이 적용할 수도 없게 됩니다.

연구팀이 해야 할 것

1. **치료 매뉴얼 작성**: 다음 내용을 상세히 문서화합니다.
    - 총 회기 수 (예: 24 회기)
    - 1 회기당 시간 (예: 30 분)
    - 총 치료 기간 (예: 12 주)
    - 회기별 세부 내용 (1-4 회기: 기초 단어, 5-8 회기: 2 단어 조합...)
2. **치료사 자격 기준**: 누가 치료를 제공할 수 있는지 명확히 정합니다.
    - 필수: 언어재활사 자격증, 최소 2 년 경력
    - 권장: 자폐 아동 치료 경험 50 시간 이상
3. **대조군 관리**: 비교 그룹이 받을 수 있는 치료와 받으면 안 되는 치료를 구체적으로 명시합니다.
    - 허용: 일반적인 작업치료, 놀이치료
    - 금지: 우리 연구와 동일한 AI 기반 언어치료

근거 수준별 연구 설계 방법 4

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 치료 프로토콜 매뉴얼 작성 (예: 회기 수, 1 회기 시간, 총 치료 기간 명시) • 치료 제공자 자격 기준 정의 (예: 면허, 경력 연수, 자격증 요건 명시) • 대조군 조건의 구체적 내용 명세 (허용·금지 치료 목록 포함)	Performance bias, Protocol deviation bias
상 (권장)	• 치료 충실도 측정 도구 적용 (체크리스트, 녹음·녹화 기반 검토) • 치료 제공자 사전 훈련 프로토콜 및 자격 인증 절차 운영 • 공동 개입 허용 범위 명시 및 기록 의무화 • 치료 충실도 정기 모니터링 (전체 회기의 20% 이상 무작위 점검 권장)	
최상 (이상)	• 치료 매뉴얼 기반 치료자 인증 시험 도입 • 세션별 실시 항목, 소요 시간, 아동 반응 전수 구조화 기록 • 충실도-효과 관계 분석 계획 수립	

최소 충족 기준 요약: 치료 프로토콜 매뉴얼화 + 치료 제공자 자격 기준 명시 + 대조군 조건 명세

자가 점검 체크리스트 4

● 기본 (필수), ● 권장 (상), ● 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
●	회기 수, 1 회기 시간, 총 치료 기간을 명시한 치료 프로토콜 매뉴얼을 작성하였다	☐	☐	☐
●	치료 제공자의 면허, 경력 연수, 자격증 요건 등 자격 기준을 정의하였다	☐	☐	☐
●	대조군에 허용되는 치료와 금지되는 치료 목록을 구체적으로 명세하였다	☐	☐	☐
●	체크리스트 또는 녹음·녹화 기반의 치료 충실도 측정 도구를 적용하였다	☐	☐	☐
●	치료 제공자 사전 훈련 프로토콜과 자격 인증 절차가 마련되어 있다	☐	☐	☐
●	공동 개입의 허용 범위를 명시하고 기록을 의무화하였다	☐	☐	☐
●	전체 회기의 20% 이상에 대해 치료 충실도 정기 모니터링을 계획하였다	☐	☐	☐
●	치료 매뉴얼 기반의 치료자 인증 시험(Certification)을 도입하였다	☐	☐	☐
●	세션별 실시 항목, 소요 시간, 아동 반응을 전수 구조화 기록하는 체계를 마련하였다	☐	☐	☐
●	충실도-효과 관계의 선량-반응 분석(Dose-response analysis) 계획을 수립하였다	☐	☐	☐

● **기본(필수) 항목 충족 현황**: 충족 ___ / 3 개

자가 점검 체크리스트 4

(부모 매개 중재 연구용 추가 체크리스트)

● 기본 (필수), ● 권장 (상), ● 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
●	부모 교육 매뉴얼을 회기별 목표, 실습 과제, 금지 사항 포함하여 작성하였다	☐	☐	☐
●	부모가 가정에서 수행할 홈미션 체크리스트를 일별/주별로 작성하였다	☐	☐	☐
●	부모 미션 수행률(예: 주당 목표 5 회 중 실제 수행 횟수)을 치료 충실도 지표로 정의하였다	☐	☐	☐
●	주간 슈퍼비전 절차를 수립하였다 (화상 또는 대면, 빈도 명시)	☐	☐	☐
●	부모가 실제로 수행하는 장면을 녹화하여 치료사가 피드백하는 절차를 마련하였다	☐	☐	☐
●	부모 Fidelity 체크리스트(핵심 행동 항목 등)를 개발하였다	☐	☐	☐
●	부모 Fidelity 최소 기준(예: 수행률 70% 이상 + 체크리스트 8/10 이상)을 사전 정의하였다	☐	☐	☐
●	부모 교육 제공자의 자격 기준(예: 부모교육 경력 3 년 이상, 관련 자격증)을 명시하였다	☐	☐	☐
●	부모 교육 제공자를 대상으로 한 사전 훈련 프로토콜 및 인증 절차를 마련하였다	☐	☐	☐
●	부모 Fidelity-효과 관계를 분석하는 선량-반응(Dose-response) 분석 계획을 수립하였다	☐	☐	☐

● **기본(필수) 항목 충족 현황**: 충족 ___ / 3 개

Domain 5. 결과 변수 선정 (Outcome Measures)

결과 변수란 무엇인가요?

결과 변수 선정은 치료 효과를 무엇으로 측정할지, 어떤 검사 도구를 사용할지 연구 시작 전에 미리 정하는 과정입니다.

왜 미리 정해야 하나요?

결과 변수를 미리 정하지 않으면, 연구가 끝난 후 여러 지표 중에서 결과가 좋게 나온 것만 골라서 보고하는 문제가 생깁니다. 예를 들어, 언어 점수는 차이가 없었는데 부모 만족도만 높았다면, "부모 만족도 향상"을 주요 성과로 포장할 수 있습니다.

또한 검증되지 않은 도구를 사용하면, 측정값 자체를 신뢰할 수 없어 신의료기술평가나 보험 급여 신청 시 근거로 인정받지 못합니다.

연구팀이 해야 할 것

1. **1 차 결과 변수 지정**: 가장 중요한 측정 지표를 딱 하나만 정합니다.
    - 예: 언어치료 연구 → 수용·표현 어휘력 검사(REVT) 점수
    - 예: 인지치료 연구 → 실행기능 평가(K-BRIEF) 총점
2. **검증된 도구 사용**: 신뢰도와 타당도가 확인된 표준화 검사를 선택합니다.
    - 국내 표준화 완료 도구 우선
    - 신뢰도 계수(Cronbach's α) 0.80 이상
    - 부록 C 의 도구 목록 참조
3. **사전 등록**: 연구 시작 전에 1 차 결과 변수를 임상시험 등록 사이트(예: ClinicalTrials.gov, CRIS)에 공개적으로 등록합니다.
4. **안전성 기준 명시**: 이상반응(부작용)과 중대이상반응(심각한 부작용) 보고 기준을 함께 정합니다.
    - 예: 문제행동 20% 이상 증가 시 이상반응으로 보고

근거 수준별 연구 설계 방법 5

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 1 차 결과 변수 단일 지정 및 사전 등록 • 표준화된 평가 도구 사용 (국내외 신뢰도·타당도 검증 도구에 한함) • 2 차 결과 변수 사전 정의 및 우선순위 명시 • 안전성 결과 변수 정의: 이상반응, 중대이상반응 보고 기준 명시 • 평가 시점 전반에 걸친 도구 동일성 유지 원칙 명시 (측정 도중 도구 교체 금지)	Selective reporting bias, Measurement bias
상 (권장)	• Minimal Clinically Important Difference 사전 정의 • 도메인별 표준화 도구 선정 근거 명시 (언어: SELSI/PRES, 행동: CBCL, 전반: K-Vineland 등) • 보호자 보고 결과 도구 병행 적용	
최상 (이상)	• Core Outcome Set 기반 결과 변수 구성 (국제 합의 기반 핵심 결과 도구 세트) • 기능적 결과 및 삶의 질 지표 병행 측정 • 장기 추적 결과 변수 사전 계획 (중재 종료 후 6 개월 이상)	

최소 충족 기준 요약: 1 차 결과 변수 사전 등록 + 표준화 도구 사용 + 평가 기준 문서화 + 이상반응/중대이상반응 정의 + 평가 중 측정 도구 교체 금지

자가 점검 체크리스트 5

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	1 차 결과 변수를 지정하고 연구 시작 전 사전 등록하였다	☐	☐	☐
🔴	신뢰도·타당도가 국내외에서 검증된 표준화 평가 도구만을 사용하기로 결정하였다	☐	☐	☐
🔴	채점 기준과 판정 기준을 포함한 평가 기준을 사전에 문서화되어 있다	☐	☐	☐
🔴	2 차 결과 변수와 우선순위를 사전에 정의하였다	☐	☐	☐
🔴	이상반응 및 중대이상반응의 보고 기준을 사전에 정의하였다	☐	☐	☐
🔴	전체 평가 시점에 걸쳐 동일한 도구를 유지하는 원칙을 명시하였다	☐	☐	☐
🟡	임상적으로 의미 있는 최소 변화량(MCID)을 사전에 정의하였다	☐	☐	☐
🟡	평가 도메인별 표준화 도구 선정 근거를 명시하였다	☐	☐	☐
🟡	보호자 보고 결과 도구를 병행 적용하기로 계획하였다	☐	☐	☐
🟢	Core Outcome Set 기반으로 결과 변수를 구성하였다	☐	☐	☐
🟢	기능적 결과 및 삶의 질 지표를 병행 측정하기로 계획하였다	☐	☐	☐
🟢	중재 종료 후 6 개월 이상의 장기 추적 결과 변수를 사전에 계획하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 6 개

자가 점검 체크리스트 5

(부모 매개 중재 연구용 추가 체크리스트)

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	아동 발달 지표를 1 차 결과 변수로 지정하였다	☐	☐	☐
🔴	양육자 역량 지표(예: 부모 기술, 양육 스트레스, 부모 효능감)를 2 차 결과 변수 또는 매개 변수로 정의하였다	☐	☐	☐
🟡	부모 보고 도구(Parent-report measures)와 관찰 기반 도구를 병행 적용하였다	☐	☐	☐
🟡	양육자 역량 → 아동 발달 경로를 검증하기 위한 매개효과 분석을 계획하였다	☐	☐	☐
🟢	부모-아동 상호작용의 질적 변화를 측정하는 영상 코딩 도구를 포함하였다	☐	☐	☐
🟢	가족 전체의 삶의 질 또는 가족 기능(Family functioning) 지표를 추가하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 2 개

Domain 6. 평가 시점 및 평가 운영 (Assessment Timing & Operations)

평가 시점이란 무엇인가요?

평가 시점 및 평가 운영은 치료 효과를 언제, 어떤 순서로 측정할지 모든 참여 기관이 똑같이 따를 수 있도록 미리 정하는 과정입니다.

왜 명확히 정해야 하나요?

평가 시점을 애매하게 정하면 기관마다 다른 시점에 측정하게 됩니다. 예를 들어, "중재 종료 후"라고만 쓰면, A 기관은 마지막 회기 바로 다음 날 측정하고 B 기관은 2 주 후에 측정할 수 있습니다. 이렇게 되면 결과를 합쳐서 분석할 수 없습니다.

또한 치료 시작 전 상태(Baseline)를 측정하지 않으면, 나중에 얼마나 좋아졌는지 변화량을 계산할 수 없습니다. "치료 후 점수 70 점"만으로는 좋아진 건지 원래 그랬던 건지 알 수 없기 때문입니다.

연구팀이 해야 할 것

1. **필수 평가 시점 3 개 포함**:
    - **T0 (Baseline)**: 치료 시작 전 (예: 1 주일 이내)
    - **T1 (Post-intervention)**: 치료 종료 직후 (예: 마지막 회기 후 1 주일 이내)
    - **T2 (Follow-up)**: 추적 평가 (예: 치료 종료 3 개월 후)
2. **구체적인 시점 명시**: 모호한 표현 대신 정확한 기간을 씁니다.
    - "중재 종료 후" X → "마지막 회기 후 4 주(±1 주)"
    - "추적 평가" X → "치료 종료 12 주 후(±2 주)"
3. **허용 오차 범위 설정**: 현실적으로 정확한 날짜에 평가하기 어려우므로 허용 범위를 함께 명시합니다.
    - 예: 12 주 ± 1 주 (11-13 주 사이 평가 가능)
4. **평가 순서 표준화**: 여러 검사를 할 때 순서를 정해둡니다.
    - 예: 1) VABS-II (30 분) → 2) REVT (20 분) → 3) CBCL (15 분)

근거 수준별 연구 설계 방법 6

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 사전 정의된 평가 시점 설정 (예: "중재 종료 후 4주" — "중재 종료 후" 등 모호한 표현 금지) • Baseline 및 Follow-up 평가 반드시 포함	Temporal bias, Deviation from intended intervention
상 (권장)	• 동일 평가 간격 및 assessment window 적용 (모든 연구대상자·기관에 동일 기준) • 평가순서 등을 포함한 표준화된 평가 운영 Standard Operating Procedure 문서화 및 전 기관 공유 • 평가 일정 이탈 기록 및 보고 의무화	
최상 (이상)	• Longitudinal repeated measures 설계 적용 (다중 시점 측정으로 효과 궤적 파악) • 평가 시점 준수율 모니터링 및 보고	

최소 충족 기준 요약: 구체적 평가 시점 사전 정의 + Baseline + Follow-up 포함

자가 점검 체크리스트 6

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	모든 연구대상자와 기관에 동일하게 적용되는 구체적인 평가 시점을 사전에 정의하였다	☐	☐	☐
🔴	치료 시작 전 기저 상태 평가를 계획에 포함하였다	☐	☐	☐
🔴	치료 종료 후 추적관찰 평가를 계획에 포함하였다	☐	☐	☐
🟡	모든 연구대상자와 기관에 동일한 평가 간격과 assessment window 를 적용하였다	☐	☐	☐
🟡	표준화된 평가 운영 SOP 를 문서화하여 전 기관에 배포하였다	☐	☐	☐
🟡	평가 일정 이탈 발생 시 기록 및 보고 절차를 수립하였다	☐	☐	☐
🟢	효과 궤적 파악을 위한 Longitudinal repeated measures 설계를 적용하였다	☐	☐	☐
🟢	연구팀 내 평가 시점 준수율을 모니터링하고 보고하는 체계를 마련하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 3 개

Domain 7. 평가자 관리 및 눈가림 (Assessor Management & Blinding)

평가자 눈가림이란 무엇인가요?

평가자 관리 및 눈가림은 아동의 상태를 측정하는 평가자가 그 아동이 치료 그룹인지 비교 그룹인지 모르는 상태에서 검사를 진행하도록 하는 과정입니다.

왜 필요한가요?

치료를 제공한 사람이 직접 평가까지 하거나, 평가자가 아동의 그룹을 미리 알고 있으면, 의도하지 않아도 무의식적으로 평가 결과가 기대하는 방향으로 나오게 됩니다.

예를 들어, "이 아이는 우리 치료를 받았으니까 좋아졌을 거야"라고 생각하면, 애매한 답변을 후하게 점수를 주거나 경계선 점수를 올림으로 처리할 수 있습니다.

실제로 **눈가림이 없는 연구는 치료 효과를 평균 40% 과대평가**한다는 연구 결과가 있습니다. 이는 결과의 신뢰성을 크게 떨어뜨립니다.

연구팀이 해야 할 것

1. **역할 분리**: 치료 제공자와 평가자를 반드시 다른 사람으로 배정합니다.
    - 치료사: 치료만 제공, 평가 참여 금지
    - 평가자: 검사만 실시, 치료 과정 모름
2. **정보 차단**: 평가자가 그룹 배정 정보를 알 수 없도록 차단합니다.
    - 평가 기록지에 그룹 정보 기재 금지
    - 아동/부모에게 "평가자 앞에서 치료 이야기 하지 말기" 안내
    - 평가실과 치료실 물리적 분리
3. **눈가림 점검(Blinding Check)**: 연구 종료 시 평가자에게 질문합니다.
    - "이 아동이 치료 그룹이었다고 생각하시나요?"
    - 정답률 50% 근처 → 눈가림 성공
    - 정답률 80% 이상 → 눈가림 실패 (정보 유출)
4. **불가능한 경우 대안**: 발달장애 연구에서 완전한 눈가림이 어려운 경우, 최소한 "평가자 눈가림"만이라도 철저히 유지합니다.

근거 수준별 연구 설계 방법 7

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 평가자와 중재 수행자(치료사) 인적 분리 • 평가자 blind 적용 (평가자에게 군 배정 정보 차단)	Measurement bias, Detection bias
상 (권장)	• 평가자 훈련 표준화 (예: 공통 훈련 프로토콜, 자격 기준 충족 확인 절차) • 평가자 간 신뢰도 측정 및 보고 • 평가자 내 신뢰도 정기 점검	
최상 (이상)	• 평가자 blind 유지 검증: Blinding Index 산출 및 보고 • 정기적 재보정 실시로 평가자 drift 방지 • 독립된 외부 평가자 활용 방안 수립 및 적용	

최소 충족 기준 요약: 평가자-치료자 인적 분리 + 평가자 blind 적용

자가 점검 체크리스트 7

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	치료를 제공하는 치료사와 효과를 평가하는 평가자를 서로 다른 사람으로 분리하였다	☐	☐	☐
🔴	평가자가 아동의 군 배정 정보(치료군/대조군)를 알 수 없도록 차단하는 절차를 수립하였다	☐	☐	☐
🟡	공통 훈련 프로토콜과 자격 기준 충족 확인 절차를 포함한 평가자 훈련 기준이 표준화되어 있다	☐	☐	☐
🟡	평가자 간 신뢰도를 측정하고 보고하는 계획을 수립하였다	☐	☐	☐
🟡	평가자 내 신뢰도를 정기적으로 점검하는 절차를 마련하였다	☐	☐	☐
🟢	눈가림 유지 성공 여부를 검증하기 위한 Blinding Index 를 산출하고 보고하기로 계획하였다	☐	☐	☐
🟢	평가자 drift 방지를 위한 정기적 재보정 일정을 수립하였다	☐	☐	☐
🟢	군 배정 정보를 모르는 독립된 외부 평가자를 활용하는 방안을 검토하고 계획하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 2 개

자가 점검 체크리스트 7

(부모 매개 중재 연구용 추가 체크리스트)

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	부모 교육을 담당하는 치료사와 아동 평가를 담당하는 평가자를 분리하였다	☐	☐	☐
🔴	부모-아동 상호작용 영상을 촬영하여 분석하는 경우, 코딩자에게 군 배정 정보를 차단하였다	☐	☐	☐
🟡	영상 파일을 비식별화(파일명에서 이름·군 정보 제거)하여 코딩자에게 제공하였다	☐	☐	☐
🟡	양육자의 심각한 정신질환, 약물/알코올 중독을 제외 기준에 포함하였다	☐	☐	☐
🟡	영상 코딩의 이중 평가(전체 영상의 20% 이상)를 계획하고 평가자 간 신뢰도를 산출하였다	☐	☐	☐
🟢	평가자가 부모와 직접 대면하지 않도록 일방향 거울 또는 별도 공간 평가 환경을 조성하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 2 개

Domain 8. 데이터 입력 및 품질 관리 (Data Collection & Quality Management)

데이터 품질 관리란 무엇인가요?

데이터 입력 및 품질 관리는 연구 과정에서 수집하는 모든 정보를 정확하고 일관된 방식으로 기록하고 관리하는 과정입니다.

왜 필요한가요?

기관마다 데이터를 다르게 기록하면, 연구가 끝나고 나서 데이터를 합치는 것 자체가 불가능해집니다.

예를 들어:

- A 기관: 성별을 "남/여"로 기록
- B 기관: 성별을 "M/F"로 기록
- C 기관: 성별을 "1/2"로 기록

→ 컴퓨터가 같은 변수로 인식하지 못해 분석 불가

또한 입력 오류나 빠진 값(결측값)을 처리하는 기준이 없으면, 연구자마다 임의로 판단하여 결과가 왜곡될 수 있습니다.

연구팀이 해야 할 것

1. **표준화된 증례기록지(CRF) 작성**: 모든 기관이 똑같은 양식을 사용합니다.
   - 변수명 통일 (예: 나이 → "age", 성별 → "sex")
   - 입력 형식 통일 (예: 날짜는 모두 "YYYY-MM-DD")
   - 코딩 규칙 통일 (예: 성별은 1=남, 2=여)
2. **입력 가이드라인 문서화**: 애매한 상황에서 어떻게 기록할지 미리 정합니다.

   변수 정의 예시:
   - "중재 참여율" = (실제 참여한 회기 수 / 전체 계획 회기 수) × 100

   결측값 처리 기준 예시:
   - 아동 거부로 검사 불가 → "9999" 입력 후 사유 기록
   - 평가 시점 미준수 → "9998" 입력 후 실제 날짜 별도 기록
3. **이중 입력 및 검증**: 중요 데이터는 두 명이 독립적으로 입력하여 일치 여부 확인합니다.
4. **정기 품질 점검**: 월 1 회 데이터를 검토하여 이상값, 오류, 결측값을 확인하고 수정합니다.
   - 예: 나이가 999 세, 점수가 -50 점 같은 명백한 오류 발견

근거 수준별 연구 설계 방법 8

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 표준화된 CRF 기반 데이터 수집 (전 기관 통일 양식 적용) • 데이터 입력 가이드라인 정의 (변수 정의, 단위, 입력 형식, 결측값 처리 기준 명문화)	Measurement bias
상 (권장)	• 데이터 검증 절차 적용 (예: 범위 오류, 논리 불일치 자동 검출) • 이중 입력 또는 주기적 Quality Control 수행 • 결측값 발생 현황 기관별 주기적 모니터링 및 보고	
최상 (이상)	• 개인정보 비식별화 프로토콜 및 데이터 보존 계획 수립	

최소 충족 기준 요약: 표준화 CRF 전 기관 적용 + 입력 가이드라인 문서화

자가 점검 체크리스트 8

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	전 기관이 동일하게 사용하는 표준화된 CRF 를 작성하였다	☐	☐	☐
🔴	변수 정의, 단위, 입력 형식, 결측값 처리 기준을 명문화한 입력 가이드라인이 마련되어 있다	☐	☐	☐
🟡	범위 오류와 논리 불일치를 자동으로 검출하는 데이터 검증 절차를 적용하였다	☐	☐	☐
🟡	이중 입력(Double data entry) 또는 주기적 QC 를 수행하는 체계를 마련하였다	☐	☐	☐
🟡	기관별 결측값 발생 현황을 주기적으로 모니터링하고 보고하는 절차를 수립하였다	☐	☐	☐
🟢	개인정보 비식별화 프로토콜과 데이터 보존 계획을 수립하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 2 개

Domain 9. 통계 분석 계획 (Statistical Analysis Plan)

통계 분석 계획이란 무엇인가요?

통계 분석 계획은 수집한 데이터를 어떤 통계 방법으로 분석하여 치료 효과를 증명할 것인지 연구 시작 전에 미리 정해두는 과정입니다.

왜 미리 정해야 하나요?

분석 방법을 미리 정하지 않으면, 연구가 끝난 후 여러 방법을 이것저것 시도해보다가 결과가 좋게 나오는 방법만 골라서 보고하는 문제가 생깁니다.

예를 들어:

- t-검정: p=0.08 (유의하지 않음)
- 비모수 검정: p=0.045 (유의함!) → 비모수 검정 결과만 보고

또한 중도 탈락한 아이들을 마음대로 제외하면, 무작위배정으로 공정하게 만들어 놓은 두 그룹의 균형이 깨집니다. 예를 들어, 치료 그룹에서 증상이 심한 아이들만 탈락했다면, 남은 아이들은 원래 증상이 가벼운 아이들이므로 결과가 좋아 보이는 것은 당연합니다.

연구팀이 해야 할 것

1. **통계 분석 계획서(SAP) 작성**: 연구 시작 전에 다음 내용을 문서화합니다.
    - 주 분석 방법 (예: 독립표본 t-검정, ANCOVA)
    - 보조 분석 방법
    - 유의수준(α): 0.05
    - 단측 vs 양측 검정

2. **ITT 원칙 명시**: 중도 탈락자도 포함하여 분석합니다.
    - ITT(Intention-to-Treat): 처음 배정된 그룹대로 모든 참여자를 분석
    - 예: 치료 그룹 배정 후 1 회만 참여하고 그만둔 아동도 치료 그룹에 포함

3. **결측값 처리 방법 명시**: 빠진 데이터를 어떻게 다룰지 미리 정합니다.
    - LOCF(Last Observation Carried Forward): 마지막 측정값을 그대로 사용
    - Multiple Imputation: 통계적 방법으로 추정값 채워넣기
    - 완전 제거: 해당 사례 제외 (비권장)

4. **사전 등록**: 작성한 SAP 를 임상시험 등록 사이트(ClinicalTrials.gov, CRIS)에 등록하여 투명성을 확보합니다.

근거 수준별 연구 설계 방법 9

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• SAP 연구 시작 전 작성 및 프로토콜 등록 • ITT 분석 원칙 준수: 배정된 군 기준 전원 분석 • 1 차 분석 방법 사전 명시 (ANCOVA, Mixed-effects model 등 선택 근거 포함) • 결측값 분석적 처리 방법 사전 정의 (예: Multiple imputation, LOCF 등)	Selective reporting bias, Attrition bias
상 (권장)	• Per-Protocol 분석 병행 및 ITT 와의 결과 비교를 통한 민감도 분석 • 하위군 분석 사전 정의 (예: 연령, 중증도, 치료 유형 등) • 다중 비교 보정 방법 명시 (예: Bonferroni, Benjamini-Hochberg 등) • 다기관 설계 시 기관 효과 보정 계획 (예: 기관을 random effect 로 처리)	
최상 (이상)	• 중간 분석 계획 및 조기 중단 기준 사전 정의 • 충실도-효과 관계 Dose-response analysis 수행 계획 • Bayesian analysis 또는 확률론적 민감도 분석을 통한 결과 견고성 검증	

최소 충족 기준 요약: SAP 사전 작성·등록 + ITT 원칙 적용 + 1 차 분석 방법 명시 + 결측값 처리 계획

자가 점검 체크리스트 9

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	연구 시작 전 SAP 를 작성하고 프로토콜에 등록하였다	☐	☐	☐
🔴	배정된 군 기준으로 모든 연구대상자를 분석하는 ITT 원칙을 적용하기로 결정하였다	☐	☐	☐
🔴	1 차 결과 변수의 분석 방법을 선택 근거와 함께 사전에 명시하였다	☐	☐	☐
🔴	결측값 처리 방법이 명시되어 있다	☐	☐	☐
🟡	ITT 분석과 PP 분석을 병행하고 두 결과를 비교하는 민감도 분석을 계획하였다	☐	☐	☐
🟡	연령, 중증도, 치료 유형 등 하위군 분석 항목을 사전에 정의하였다	☐	☐	☐
🟡	다중 비교 보정 방법을 명시하였다	☐	☐	☐
🟡	다기관 설계 시 기관 효과를 처리하는 계획을 수립하였다	☐	☐	☐
🟢	중간 분석 계획과 조기 중단 기준을 사전에 정의하였다	☐	☐	☐
🟢	충실도-효과 관계의 Dose-response analysis 계획을 수립하였다	☐	☐	☐
🟢	Bayesian analysis 또는 확률론적 민감도 분석을 통한 결과 견고성 검증을 계획하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 4 개

Domain 10. 보건경제성 분석 (Health Economic Analysis)

보건경제성 분석이란 무엇인가요?

보건경제성 분석은 우리 치료법이 효과가 있을 뿐만 아니라, 그 효과를 얻는 데 드는 비용이 사회적으로 받아들일 만한 수준인지 따져보는 과정입니다.

왜 필요한가요?

임상 효과가 입증되어도 비용 측면의 근거가 없으면, 건강보험심사평가원(HIRA)의 급여 심사에서 "경제성이 부족하다"는 이유로 거절될 수 있습니다. 아무리 좋은 치료법이라도 너무 비싸면 보험 적용이 안 되는 것입니다.

또한 치료비만 계산하면 발달장애 가정이 실제로 부담하는 비용의 일부만 반영됩니다. 병원까지 가는 교통비, 보호자가 일을 쉬는 비용, 형제자매를 맡기는 비용 등이 모두 빠지게 됩니다.

연구팀이 해야 할 것

1. **비용 수집 항목을 CRF 에 포함**: 연구 중 발생하는 모든 비용을 기록합니다. **직접 의료비용**:
    - 치료비 (회기당 비용 × 횟수)
    - 검사비 (VABS-II, REVT 등)
    - 약제비 (해당 시)
    - 교통비 (왕복 거리 × 유가)

    간접 비용 (권장):
    - 보호자 시간 비용 (시급 × 소요 시간)
    - 생산성 손실 (조퇴, 결근)
    - 아이 돌봄 비용

2. **분석 관점 명시**: 누구 입장에서 비용을 계산할지 정합니다.
    - **보험자 관점**: 건강보험공단이 지불하는 비용만
    - **환자 관점**: 가족이 직접 부담하는 본인부담금
    - **사회적 관점**: 위 모두 포함 (권장)

3. **비용-효과 분석**: 비용 대비 얼마나 효과적인지 계산합니다.
    - ICER(증분비용효과비) = (치료 그룹 비용 - 비교 그룹 비용) / (치료 그룹 효과 - 비교 그룹 효과)
    - 예: VABS-II 1 점 향상에 10 만원 소요

4. **급여 기준 확인**: HIRA 가 제시하는 비용-효과성 기준치와 비교합니다.

근거 수준별 연구 설계 방법 10

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 직접 의료비용 수집 항목 사전 정의 (예: 치료비, 검사비, 교통비 등) • 비용 수집 기간 및 방법 명시 (CRF 내 비용 기록 항목 포함) • 분석 관점 사전 명시: 지불자 관점 최소 적용	Measurement bias
기본 (필수)	• QALY 산출을 위한 효용값 측정 도구 적용 (EQ-5D-Y 또는 HUI3) • 보호자 생산성 손실 등 간접 비용 측정 계획 수립	

최소 충족 기준 요약: 직접 의료비용 수집 계획 수립 + 분석 관점 사전 명시

자가 점검 체크리스트 10

● 기본 (필수), ● 권장 (상), ● 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
●	치료비, 검사비, 교통비 등 직접 의료비용 수집 항목을 사전에 정의하였다	☐	☐	☐
●	비용 수집 기간과 방법을 명시하고 CRF 에 비용 기록 항목이 포함되어 있다	☐	☐	☐
●	지불자 관점을 포함한 분석 관점을 사전에 명시하였다	☐	☐	☐
●	QALY 산출을 위한 효용값 측정 도구(EQ-5D-Y 또는 HUI3)를 적용하기로 계획하였다	☐	☐	☐
●	보호자 생산성 손실을 포함한 간접 비용을 측정하는 방법을 계획하였다	☐	☐	☐

● **기본(필수) 항목 충족 현황**: 충족 ___ / 3 개

Domain 11. 윤리 및 규제 준수 (Ethics & Regulatory Compliance)

윤리 및 규제 준수란 무엇인가요?

윤리 및 규제 준수는 연구에 참여하는 아동과 가족을 보호하고, 연구가 국내외에서 인정하는 기준에 따라 적법하게 진행되고 있음을 증명하는 과정입니다.

왜 필요한가요?

IRB(기관생명윤리위원회) 승인이나 임상연구 사전 등록 없이 진행된 연구는, 결과가 아무리 좋아도 나중에 체계적 문헌고찰이나 메타분석에서 아예 제외됩니다. 즉, 연구 자체가 근거로 인정받지 못합니다.

또한 발달장애 아동은 스스로 참여 여부를 충분히 판단하기 어렵습니다. 그렇다고 부모 동의만 받으면 되는 것도 아닙니다. 아동도 자신이 이해할 수 있는 수준에서 설명을 듣고 거부할 권리가 있습니다.

연구팀이 해야 할 것

1. **IRB 승인 획득**: 연구 시작 전 반드시 기관생명윤리위원회의 승인을 받습니다.
    - 제출 서류: 연구계획서, 동의서, 설문지, CRF 등
    - 승인 후에도 연간 지속심의 필요
    - 프로토콜 변경 시 수정 승인 필수
2. **임상연구 사전 등록**: 연구 시작 전 공개 등록 시스템에 등록합니다.
    - 국내: CRIS (https://cris.nih.go.kr)
    - 국제: ClinicalTrials.gov
    - 등록 내용: 연구 목적, 대상자 수, 1 차 결과 변수, 평가 시점
3. **보호자 서면 동의**: 부모 또는 법정대리인에게 충분히 설명하고 동의를 받습니다.
    - 동의서 필수 포함 내용:
        - ✓ 연구 목적과 방법
        - ✓ 예상되는 이익과 위험
        - ✓ 참여 거부 및 중도 철회 권리
        - ✓ 개인정보 보호 방침
4. **아동 동의(Assent) 절차**: 아동의 연령과 이해 수준에 따라 추가로 동의를 구합니다.
    - 만 7 세 이상: 아동용 설명문(쉬운 언어, 그림 포함) 제공 후 구두 동의
    - 만 5-6 세: 그림카드로 설명 후 "하기 싫으면 안 해도 돼" 명확히 전달
    - 아동이 거부하면 부모 동의와 무관하게 참여 불가
5. **안전성 모니터링**: 중대한 이상반응(예: 자해 행동 증가 등) 발생 시 즉시 IRB 보고

근거 수준별 연구 설계 방법 11

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• IRB 승인 취득 (연구 시작 전 필수, 다기관 시 각 기관 또는 중앙 IRB) • 임상연구 공식 등록: CRIS(예: 임상연구정보서비스, WHO 인증) 또는 ClinicalTrials.gov • 소아 대상 대리동의(Parental consent) + Assent 절차 수립 • 이해충돌 선언 및 관리 계획 문서화	Ethical bias, Reporting bias
상 (권장)	• CONSORT 2025 및 SPIRIT 2013 기반 연구 프로토콜 작성 • GCP 준수 선언 및 관련 교육 이수 기록 • 연구 현장 모니터링 계획 수립 (방문 주기 및 점검 항목 명시) • 경제적 어려움으로 진단 지연 아동 우선 참여 • 연구 종료 후 필요 시 치료 연계 지원	
최상 (이상)	• Data Safety Monitoring Board 구성 및 운영 (독립적 안전성 모니터링 위원회) • 중간 분석 결과에 따른 조기 중단 윤리 기준 사전 정의 • 외부 감사 수감에 대비한 내부 자료 관리 체계 구축	

> **최소 충족 기준 요약:** IRB 승인 + CRIS 임상연구 등록 + 대리동의·Assent 절차 + 이해충돌 선언

자가 점검 체크리스트 11

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	연구 시작 전 IRB 승인을 취득하였다 (다기관 시 각 기관 또는 중앙 IRB)	☐	☐	☐
🔴	CRIS(임상연구정보서비스) 또는 ClinicalTrials.gov 에 임상연구를 사전 등록하였다	☐	☐	☐
🔴	보호자 서면 동의와 아동 본인 동의 절차를 수립하였다	☐	☐	☐
🔴	이해충돌 선언과 관리 계획을 문서화하였다	☐	☐	☐
🟡	CONSORT 2025 및 SPIRIT 2013 기준에 따라 연구 프로토콜을 작성하였다	☐	☐	☐
🟡	GCP 준수를 선언하고 관련 교육 이수 기록을 갖추고 있다	☐	☐	☐
🟡	현장 방문 모니터링 주기와 점검 항목을 포함한 모니터링 계획을 수립하였다	☐	☐	☐
🟢	독립적 안전성 모니터링을 위한 Data Safety Monitoring Board 를 구성하였다	☐	☐	☐
🟢	중간 분석 결과에 따른 조기 중단 윤리 기준을 사전에 정의하였다	☐	☐	☐
🟢	외부 감사 수감에 대비하여 연구 기록 및 자료를 체계적으로 관리하는 내부 절차를 마련하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 4 개

Domain 12. 일반화 가능성 및 정책 연계 전략

(Generalizability & Policy Translation)

일반화 가능성이란 무엇인가요?

일반화 가능성 및 정책 연계 전략은 우리 연구 결과가 특정 병원이나 특수한 상황에만 해당하는 것이 아니라, 전국 어디서나 다양한 아동에게 적용될 수 있는지 확인하고, 그 근거를 정책이나 보험 급여로 연결하기 위한 전략을 세우는 과정입니다.

왜 필요한가요?

참여 기관이나 대상 아동이 너무 제한적이면, "이 결과는 특별한 경우에만 해당한다"는 이유로 건강보험 급여 적용 근거로 인정받기 어렵습니다.

예를 들어:

- 서울 대학병원 1 곳에서만 연구
- 고소득층 가정의 경증 아동만 참여 → HIRA: "전국의 모든 발달장애 아동에게 효과가 있다고 보기 어렵다"

즉, 연구의 **내적 타당도**(이 연구 안에서는 효과가 있다)는 높아도, **외적 타당도**(다른 곳에서도 효과가 있다)가 낮으면 정책적으로 채택되기 어렵습니다.

연구팀이 해야 할 것

1. **다양성 확보**: 참여자 특성을 다양하게 구성합니다. **지역적 다양성**:
    - 수도권 + 지방 도시 + 농촌 지역
    - 최소 3 개 이상 지역에서 모집

    사회경제적 다양성:
    - 가구 소득 수준: 저소득층, 중산층, 고소득층 포함
    - 의료보험 유형: 건강보험 + 의료급여
    - 부모 학력: 고졸 이하 ~ 대졸 이상

    증상 중증도 다양성:
    - 경증 ~ 중등도 아동 모두 포함
    - 단, 안전을 고려하여 극단적 중증은 제외 가능
2. **참여자 특성 상세 기술**: 논문 작성 시 Table 1 에 다음 정보를 포함합니다.
    - 나이, 성별, 진단명
    - 가구 소득, 부모 학력
    - 거주 지역 (수도권/지방)

- 증상 중증도 분포

3. **내적 타당도 vs 외적 타당도 구분**: **내적 타당도** (연구의 엄격성):
    - 무작위배정 준수
    - 평가자 눈가림
    - ITT 분석 → "이 연구 안에서 효과가 진짜다"

   외적 타당도 (적용 범위):
    - 다양한 지역, 소득, 중증도
    - 실제 임상 환경 반영 → "다른 곳에서도 효과가 있을 것이다"
4. **정책 연계 전략 수립**: 연구 결과를 정책으로 연결하는 방안을 미리 계획합니다.
    - HIRA 급여 신청 로드맵 작성
    - 보건복지부 정책 제안 준비
    - 관련 학회·협회와 협력 체계 구축

근거 수준별 연구 설계 방법 12

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 참여 아동의 사회경제적, 지역적 다양성 기술 • 결과 보고 시 내적 타당도와 외적 타당도 구분하여 기술 • SCI/SCIE 급 국제학술지 게재 계획 수립 (GRADE 근거의 국제적 신뢰도 확보)	Generalizability bias, Applicability bias
상 (권장)	• 연구 참여 기관의 유형 다양성 확보 (예: 병원, 발달센터, 클리닉 혼합) • PRECIS-2 도구 적용: 연구 설계의 실용성 수준 정량적 제시	

> **최소 충족 기준 요약:** 참여 집단 특성 기술 + 내·외적 타당도 구분 보고 + SCI/SSCI 게재 계획

자가 점검 체크리스트 12

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	참여 아동, 양육자 및 가정의 지역적·사회경제적 다양성을 기술하는 계획을 수립하였다	☐	☐	☐
🔴	결과 보고 시 내적 타당도와 외적 타당도를 구분하여 기술하는 방안이 계획되어 있다	☐	☐	☐
🔴	SCI/SCIE 급 국제 학술지 게재 계획을 수립하였다	☐	☐	☐
🟡	병원·발달센터·클리닉 등 다양한 유형의 기관이 연구에 참여하도록 구성하였다	☐	☐	☐
🟡	PRECIS-2 도구를 활용하여 연구 설계의 실용성 수준을 정량적으로 제시하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 3 개

전체 도메인 통합 요약표

순서	Domain	최소 충족 기준	미충족 시 위험	연구 단계
1	**연구 설계 유형**	RCT 기반 + 기관 수 + 연구 기간 명시	근거 수준 하향	**설계**
2	**연구대상자 선정 및 무작위배정**	포함/제외 기준 + 비교군 + random generation + allocation concealment	High RoB	**설계**
3	**표본 크기 산출**	Power analysis (a=0.05, power≥80%) + 효과 크기 근거 + 최소 160 명 (각 80 명) + 탈락률 반영	근거 불충분 판정	**설계**
4	**중재 프로토콜 표준화**	치료 매뉴얼화 + 제공자 자격 기준 + 대조군 조건 명세	Performance bias	**설계**
5	**결과 변수 선정**	1 차 변수 사전 등록 + 표준화 도구 + 평가 기준 문서화 + AE/SAE 정의	Selective reporting bias	**설계**
6	**평가 시점 및 운영**	구체적 시점 사전 정의 + Baseline + F/U 포함	Temporal bias	**실행**
7	**평가자 관리 및 눈가림**	평가자·치료자 분리 + 평가자 blind 적용	Detection bias	**실행**
8	**데이터 입력 및 품질관리**	표준화 CRF 적용 + 입력 가이드라인 문서화	Measurement bias	**실행**
9	**통계 분석 계획**	SAP 사전 등록 + ITT 원칙 + 1 차 분석 방법 명시 + 결측값 처리 계획	Selective reporting bias	**분석**
10	**보건경제성 분석**	직접 의료비용 수집 계획 + 분석 관점 명시	급여 심사 탈락	**분석**
11	**윤리 및 규제 준수**	IRB 승인 + CRIS 등록 + 대리동의 + 이해충돌 선언	연구 적격성 박탈	**전 단계**
12	**일반화 및 정책 연계**	집단 특성 기술 + 내·외적 타당도 구분 보고 + SCI/SSCI 게재 계획	Applicability 하향	**보고**

핵심원칙

Domain 1~5 는 연구 시작 전 설계 단계에서 모두 확정되어야 하며, Domain 6~8 은 연구 수행 중 실행 기준, Domain 9~10 은 분석 단계, Domain 11 은 전 단계에 걸친 필수 요건, Domain 12 는 결과 보고 및 정책 연계 단계에 해당합니다.

부록 A. 진단검사 정확도 연구 가이드라인

진단검사 정확도(Diagnostic Test Accuracy, DTA) 연구는 새로운 진단검사(Index test)가 가장 확실한 진단 방법(Gold standard)과 비교하여 질환을 얼마나 정확하게 구분해내는지를 평가하는 연구로, 본 부록은 DTA 연구를 수행하는 팀을 위한 설계 기준을 제시합니다.

1. DTA 연구의 기본 구조

DTA 연구란 무엇인가요?

DTA(Diagnostic Test Accuracy) 연구는 우리가 새로 개발한 진단 도구(Index test)가 질환을 얼마나 정확하게 찾아내는지 평가하는 연구입니다. 예를 들어, 자폐증을 조기에 선별하는 새로운 앱이나 설문지의 정확도를 확인하는 것입니다.

DTA 연구의 핵심 구조

모든 연구 참여자는 다음 두 가지를 **모두** 받습니다:

1. **Index test (평가 대상 검사)**
    - 우리가 개발한 새로운 선별 도구
    - 예: 스마트폰 앱 기반 자폐증 선별 검사 (5 번 팀)
    - 빠르고 간편하지만 정확도는 검증되지 않음
2. **Reference standard (참고 표준)**
    - 가장 확실한 진단 방법 ("정답")
    - 예: 전문가의 DSM-5 진단 + ADOS-2
    - 정확하지만 시간과 비용이 많이 듦

왜 모두가 두 검사를 받나요?

Index test 의 결과와 Reference standard 의 결과를 비교해야 정확도를 계산할 수 있기 때문입니다.

예시: 100 명 참여

- 100 명 모두 → Index test (우리 앱) 실시
- 100 명 모두 → Reference standard (전문가 진단) 실시

결과 비교:

- Index test "양성" + Reference standard "ASD" → 진양성 (맞힘!)
- Index test "양성" + Reference standard "정상" → 위양성 (틀림)
- Index test "음성" + Reference standard "ASD" → 위음성 (놓침!)
- Index test "음성" + Reference standard "정상" → 진음성 (맞힘!)

연구팀이 주의할 점

1. **검사 순서**: Index test 를 먼저 하고, Reference standard 는 나중에 합니다.
    - 이유: 전문가 진단을 먼저 하면 Index test 결과가 영향받을 수 있음
2. **눈가림**: Index test 평가자와 Reference standard 평가자는 서로의 결과를 모르게 합니다.
    - 이유: "이 아이는 앱에서 양성 나왔으니 진단 내려야지" 같은 편향 방지
3. **동일 시점 평가**: 두 검사를 비슷한 시기에 실시합니다.
    - 권장: 2 주 이내
    - 이유: 시간 차이가 크면 아동 상태가 바뀔 수 있음

1.1 DTA 연구 vs 치료 중재 RCT

항목	치료 중재 RCT	DTA 연구
연구 질문	치료가 효과가 있는가?	진단검사가 정확한가?
Intervention	치료 (약물, 재활 등)	진단검사 (Index test)
Comparator	대조군 (위약, 표준치료 등)	표준진단검사(Reference test - Gold standard)
Outcome	치료 효과 (발달지표 등)	진단적 정확도 (민감도, 특이도, AUROC 등)
연구 설계	RCT (무작위배정 필수)	Cross-sectional or Cohort (눈가림 필수)
대상자 배정	치료군 vs 대조군	모든 연구대상자가 index test, reference test 모두 받음
눈가림	평가자가 군 배정 모름	Index test 판독자가 reference test 결과 모름. (반대의 경우도 동일함)
통계 분석	평균 차이 검정	민감도, 특이도, AUROC, 우도비 등

1.2 DTA 연구의 핵심 구성 요소

- Index test: 평가 대상인 새로운 진단검사
- Reference standard: 질환 유무를 판정하는 참조표준 (예: DSM-5 기반 임상 진단)
- Target condition: 진단하려는 질환 (예: 자폐스펙트럼장애, 발달지연)
- Target population: 검사를 적용할 대상 인구집단

2. 발달장애 DTA 연구의 특수성

2.1 Reference standard 의 불완전성 문제

일반적인 DTA 연구에서 Reference standard 는 "완벽한 Gold standard"를 가정함:

- 병리조직검사 (조직을 직접 확인)
- 유전자검사 (객관적 검사 결과)
- 영상의학 검사 + 장기 추적관찰

하지만 발달장애 진단의 Reference standard 는 이에 비해 불완전함:

문제점	설명	영향
행동관찰 기반	발달장애는 생물학적 표지자가 없으며 아동의 행동 발달을 관찰하여 진단함	평가자의 주관이 개입될 여지가 큼
진단기준의 모호성	DSM-5 진단기준이 있으나 “사회적 의사소통의 지속적 결함” 등과 같이 기준을 정량화하기 어려움	동일 아동을 다른 전문의가 보면 진단이 달라질 수 있음
평가도구의 한계	ADOS-2, ADI-R 같은 표준화 도구도 관찰자의 훈련도, 경험에 따라 점수가 달라짐	평가자 간 신뢰도가 완벽하지 않음. (Kappa 0.7-0.8 수준)
발달적 변이성	영유아 발달은 개인차가 크고 시기에 따라 급변하므로 “정상/비정상” 경계(cutoff)가 불분명함	동일 아동의 다른 시점의 평가 결과가 바뀔 수 있음
공존질환	ASD. 언어지연, 지적장애 등이 중복되는 경우가 많아 명확한 범주화가 어려울 수 있음	Reference standard 자체가 “진실”이 아닐 수 있음

2.2 Imperfect Reference Standard 에 대한 대응 전략

▣ 전략 1: Reference Standard 의 신뢰도 강화 (기본 필수)

조치	구체적 방법	목표
다학제 진단팀 구성	소아청소년정신과 전문의 1 인, 임상발달심리사 1 인을 포함한 3 인 이상의 전문가로 구성된 다학제 진단팀 구성	최종 진단의 신뢰도 향상
다중 표준화 도구 통합 사용	예) DSM-5 +ADOS-2+ADI-R 복합 사용	주관성 최소화
반복 측정	T0 시점 - Index test + Reference test T1 시점 (6 개월 후) - Index test + Reference test	Reference test 타당도 강화, Index test 예측 타당도 평가
평가자 훈련	평가자 대상 사전 훈련 및 인증	판정 기준 통일
일치도 측정	진단팀이 2 개 이상일 경우 Fleiss' Kappa, Krippendorff's alpha 등 산출	Reference standard 신뢰도 정량화

▣ 전략 2: 민감도 분석 (기본 필수)

Reference standard 의 정의를 바꿔가며 Index test 성능 변화 확인:

예시: 시나리오 1: 엄격한 진단 (ADOS + ADI-R 모두 cut-off 초과)
시나리오 2: 완화된 진단 (ADOS 또는 ADI-R 중 하나만 초과)
시나리오 3: 임상 판단 중심 (도구 결과와 무관하게 전문의 진단)

2.3 발달장애 DTA 연구 특유의 비뚤림 (기본 필수)

Bias 유형	정의	예방 방법
Spectrum bias	연구 대상자의 질환 중증도 분포가 실제 임상과 다름	다양한 중증도 포함, 지역사회 기반 모집
Verification bias	Index test 양성자만 Reference standard 받음	전 대상자가 Reference standard 받도록 설계 (Complete verification)
Differential verification bias	Index test 결과에 따라 다른 Reference standard 적용	모든 대상자에게 동일한 Reference standard 적용
Incorporation bias	Index test 결과가 Reference standard 판정에 포함됨	Reference standard 판정 시 Index test 결과 블라인드
Test review bias	Index test 판독 시 Reference standard 결과를 앎	Index test 결과 먼저 고정, 이후 Reference standard 수행
Diagnostic review bias	Reference standard 판정 시 Index test 결과를 앎	전문의는 Index test 결과를 모르는 상태에서 진단

3. DTA 연구 설계 기준 (SPACE 2 적용)

Domain 1. 연구 설계 유형 (Study Design)

DTA 연구는 새로운 진단검사(Index test)가 질환을 얼마나 정확하게 구분해내는지를 평가합니다. 치료 중재 RCT 와 달리 제비뽑기식 그룹 나누기가 필요 없으며, 모든 참여자가 Index test 와 Reference standard 를 모두 받습니다.

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• Prospective cohort study - T0: index test + reference test - T1(6 개월 후): index test + reference test 재검사 • 모든 연구대상자가 Index test 와 Reference standard 모두 수행 (Complete verification) • Index test 와 Reference standard 수행 순서 및 블라인드 유지 방법 사전 명시	Verification bias, Test review bias, Diagnostic review bias
상 (권장)	• 다기관 설계: 최소 3 개 이상 기관 참여	
최상 (이상)	• 외부 검증(External validation): 독립적인 외부 데이터셋에서 Index test 성능 재검증 • 시간적 검증(Temporal validation): 시간 간격을 두고 수집한 데이터에서 성능 확인	

최소 충족 기준 요약: Prospective cohort study 설계 + Complete verification + 블라인드 유지 계획

Domain 2. 대상자 선정 (Participant Selection)

DTA 연구에서는 대상 집단을 명확히 정의하고, 다양한 질환 중증도를 포함하여 스펙트럼 편향(특정 정도의 환자만 모집되는 문제)을 막아야 합니다. 발달장애 DTA 연구에서는 증상이 가벼운 아동부터 심한 아동까지 고르게 포함하는 것이 특히 중요합니다.

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• Target population 명확히 정의 • 제외 기준: 명확한 유전/대사/뇌병변 질환, Index test 수행 불가 장애, **이미 확진받은 경우** • 지역사회 기반 모집: 지역사회 어린이집 연계 모집으로 경증 사례 확보	Spectrum bias, Selection bias
최상 (이상)	• 모집 및 배정 과정의 사전 정의와 재현 가능성 확보 • 무작위배정 전 과정 문서화 (CONSORT 흐름도 포함)	

> **최소 충족 기준 요약**: Target population 정의 + 포함/제외 기준 명문화 + 다양한 중증도 포함

Domain 3. 표본 크기 산출 (Sample Size Calculation)

DTA 연구의 표본 크기는 목표로 하는 민감도·특이도와 예상되는 유병률을 기준으로 계산합니다. RCT 의 치료 효과 크기를 바탕으로 계산하는 방법과는 다른 공식을 사용합니다.

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 목표 민감도·특이도 설정 (예: Se ≥80%, Sp ≥80%) • 연구 표본에서 예상되는 발달장애 유병률 고려 • DTA 표본 크기 공식 적용: - 질환자 수: $n_{cases} = (Z^2\alpha/2 \times Se \times (1-Se)) / d^2$ - 총 표본 수 = n_{cases} / 예상 유병률 • 예상 탈락률 반영한 초기 모집 목표 수 설정 (통상 15~20% 추가 모집) • SPACE 2 최소 기준: 전체 160 명 이상 (case/control 각각 80 명 이상)	Imprecision
상 (권장)	• 층화변수별 하위군 분석 고려한 추가 표본 크기 검토	
최상 (이상)	• 외부 검증용 독립 데이터셋 확보: 다른 기관 또는 다른 시기의 데이터로 성능 재검증	

최소 충족 기준 요약: DTA 표본 크기 공식 적용 + 목표 Se/Sp 명시 + 최소 160 명 확보

Domain 4-1. Index Test 수행 및 관리

Index test 는 정확도를 확인하려는 새로운 진단검사를 말합니다. 통일된 검사 방법과 눈가림 유지가 공정한 결과를 얻기 위해 꼭 필요합니다.

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• Index test 명확히 정의: 검사 방법, 장비, 소프트웨어 버전, **Cut-off 값 명시** • 표준화 프로토콜 수립: 검사 수행 조건 명시, Index test 시행 후 reference test 시행 • 결과 보고 형식: 이분형(양성/음성) + 연속형(확률값/점수) 모두 수집하여 ROC curve 작성 • 품질 관리 기준: 부적절한 결과 제외 기준 사전 정의 • **블라인드 유지:** Index test 판독자는 Reference standard 결과를 모르는 상태에서 판정	Test review bias, Incorporation bias
상 (권장)	• 검사자 자격 및 훈련: Index test 수행자 자격 기준 명시, 사전 훈련 프로토콜 • 메타데이터 수집: 검사 일시, 장비, 영상 품질 등급 기록	
최상 (이상)	• 영상 품질 평가자 간 일치도 측정: 품질 판정의 일관성 확인 • 다양한 촬영 장비에서 성능 평가: Index test 의 일반화 가능성 검증 • 영상 품질에 따른 성능 차이 분석: 품질 등급별 민감도·특이도 변화	

> **최소 충족 기준 요약**: Index test 명확히 정의 + 표준화 프로토콜/결과 보고형식 수립 + 품질 관리 기준 + 블라인드 유지

Domain 4-2. Reference Standard 정의 및 수행

Reference standard 는 질병이 있는지 없는지를 판단하는 기준("정답")입니다. 발달장애는 행동을 관찰하여 진단하기 때문에 Reference standard 가 완벽하지 않다는 점을 인정하고 신뢰도를 높여야 합니다.

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• Reference Standard 명확히 정의: 임상 진단 (DSM-5 진단기준 적용) 필수, T1 시점 진단 적용 • 3 인 이상 다학제 진단팀 구성 - 소아청소년정신과 전문의 1 인, 임상발달심리사 1 인을 포함한 3 인 이상의 전문가로 구성 - 전문가 자격 기준: 발달장애 진단 경력 10 년 이상 • 다중 표준화 도구 통합 사용 (예: 임상 진단 + 표준화 도구 (ADOS-2, ADI-R) + 발달력 검토 통합) • 블라인드 유지: 다학제 진단팀은 Index test 결과를 모르는 상태에서 진단 • 다학제 진단팀 간 일치도 측정 및 보고 (2 팀 이상 다학제 진단팀 운영시)	Diagnostic review bias, Imperfect reference standard
상 (권장)	• ADOS-2 공인 평가자 자격 보유 전문의 포함	
최상 (이상)	• Latent Class Analysis: Reference standard 없이 통계적 모델로 "진짜 질환 상태" 추정 (고급 통계 기법)	

최소 충족 기준 요약: Reference standard 정의 (임상 진단 필수) + 다학제 진단팀 구성 + + 다중 표준화 도구 통합 사용 + 블라인드 유지 + 진단팀 간 일치도 측정·보고 (해당시)

Domain 5. 결과 변수 선정 (Outcome Measures)

DTA 연구의 1 차 결과는 진단이 얼마나 정확한지 나타내는 지표(민감도, 특이도, AUROC 등)이며, 2 차 결과로 Reference standard 의 신뢰도를 함께 보고합니다.

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
1 차 결과 변수 (Index test 진단 정확도)		
기본 (필수)	• 1 차 결과 변수 (진단 정확도 지표): - 민감도 (Sensitivity): TP / (TP + FN), 목표 ≥80% - 특이도 (Specificity): TN / (TN + FP), 목표 ≥80% - AUROC: ROC curve 아래 면적, 목표 ≥0.8 - Kappa coefficient (Index test-Reference standard): 일치도 • 모든 지표에 95% CI 제시 • 2×2 분할표 작성 (TP, FP, TN, FN)	Selective reporting bias
상 (권장)	• 반복측정으로 인한 결과 변수: 진단 안정성, 예측 민감도, 예측 특이도, test-retest reliability	
최상 (이상)	• 하위군별 분석: 연령별, 중증도별, 질환 유형별 민감도·특이도 • 임상적 유용성 지표: Decision curve analysis (Net benefit), Calibration plot	
2 차 결과 변수 (Reference Standard 신뢰도)		
기본 (필수)	• 전문의 간 일치도 (Inter-rater reliability): - Fleiss' Kappa (3 명 이상), Cohen's Kappa (2 명) - 95% CI 제시 - 목적: Reference standard 가 얼마나 신뢰할 수 있는지 보여줌	

최소 충족 기준 요약: 1 차 결과 변수 사전 등록 + 2×2 분할표 작성 + Reference standard Kappa + 진단팀 간 Kappa

Domain 6. 평가 시점 및 평가 운영 (Assessment timing & operations)

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 사전 정의된 평가 시점 설정 • Index test 후 Reference test 시행 • T0 및 T1 평가 반드시 포함 - T0: index test + reference test - T1(6 개월 후): index test + reference test 재검사	Temporal bias, Deviation from intended intervention
상 (권장)	• 동일 평가 간격 및 assessment window 적용 (모든 연구대상자·기관에 동일 기준) • 표준화된 평가 운영 Standard Operating Procedure 문서화 및 전 기관 공유 • 평가 일정 이탈 기록 및 보고 의무화	
최상 (이상)	• 평가 시점 준수율 모니터링 및 보고	

최소 충족 기준 요약: 구체적 평가 시점 사전 정의 + Index test 후 reference test 순서 지정 + T0, T1 평가 포함

Domain 7. 평가자/진단팀 관리 및 눈가림

DTA 연구 특유의 결과 왜곡을 QUADAS-2 도구로 평가하여 막습니다. 발달장애 DTA 는 완벽하지 않은 Reference standard 로 인한 추가 위험이 있으므로 이를 고려해야 합니다.

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 모든 대상자에게 Index test 양성/음성 여부와 무관하게 동일한 Reference standard 적용 • Index test 평가자에게 reference test 결과 blind 적용 • 다학제 진단팀에게 index test 결과 blind 적용 • Index test 수행 후 reference test 수행 • Index test 결과가 reference test 판정에 포함되지 않도록 차단	Verification bias, Diagnostic review bias, Test review bias
상 (권장)	• 지역사회 기반 모집으로 경증 사례 확보	
최상 (이상)	• 평가자-진단팀 blind 유지 검증: Blinding Index 산출 및 보고 • 독립된 외부 진단팀 활용 방안 수립 및 적용	

> **최소 충족 기준 요약**: 모든 대상자 index test-reference test 수행 + Index test 평가자-다학제 진단팀 간 blind 적용 + Index test 후 reference test 수행 + reference test 결과에 index test 결과 포함 차단

Domain 8. 데이터 입력 및 품질 관리 (Data collection and quality management)

DTA 연구는 STARD 2015 체크리스트에 따라 보고해야 하며, 특히 발달장애 DTA 의 특성상 완벽하지 않은 Reference standard 의 한계를 솔직하게 밝혀야 합니다.

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 표준화된 CRF 기반 데이터 수집 (전 기관 통일 양식 적용) • 데이터 입력 가이드라인 정의 (변수 정의, 단위, 입력 형식, 결측값 처리 기준 명문화) • STARD 2015 체크리스트 준수 (flow diagram 작성 포함)	Measurement bias
상 (권장)	• 데이터 검증 절차 적용 (예: 범위 오류, 논리 불일치 자동 검출) • 이중 입력 또는 주기적 Quality Control 수행 • 결측값 발생 현황 기관별 주기적 모니터링 및 보고	
최상 (이상)	• 개인정보 비식별화 프로토콜 및 데이터 보존 계획 수립	

> **최소 충족 기준 요약**: 표준화 CRF 전 기관 적용 + 입력 가이드라인 문서화 + STARD 준수

Domain 9. 통계 분석 계획 (Statistical Analysis Plan)

DTA 연구의 통계 분석은 2×2 분할표를 바탕으로 진단 정확도 지표를 계산하는 것이 핵심입니다.

근거 수준	연구 설계 방법	통제 대상 비뚤림 (RoB)
기본 (필수)	• 2×2 분할표 작성 • 정확도 분석 변수: 민감도, 특이도, 95% CI 산출 • ROC curve 작성 (Index test 가 연속형 결과 제공 시): • Kappa coefficient: - Index test-Reference standard 간 (주요 분석) - 다학제 진단팀 간 (Reference standard 신뢰도) • 민감도 분석: Reference standard 정의 변경 시 Index test 성능 변화	Selective reporting bias, Attrition bias
상 (권장)	• 하위군 분석 사전 정의 (예: 연령, 중증도, 치료 유형 등) - 연령별, 중증도별, 질환 유형별 민감도·특이도 - Forest plot 으로 시각화	
최상 (이상)	• Calibration plot: AI 예측 확률과 실제 발생률 일치도 평가, Hosmer-Lemeshow test • Decision curve analysis: 임상적 의사결정에서의 net benefit	

> **최소 충족 기준 요약**: 2×2 분할표 + Se/Sp (95% CI) + ROC curve & AUROC + Kappa 산출 + 민감도 분석

Domain 10 ~12

Domain 10. 보건경제성 분석 ~ Domain 12. 일반화 가능성 및 정책 연계 전략의 내용과 자가 점검 체크리스트는 중재연구와 동일한 기준으로 적용됩니다.

전체 Domain 통합 요약표 (DTA 연구)

순서	Domain	최소 충족 기준	미충족 시 위험	연구 단계
1	**연구 설계 유형**	Prospective cohort + T0·T1 반복측정 + Complete verification + Blind	근거 수준 하향	**설계**
2	**연구대상자 선정**	Target population 정의 + 포함/제외 기준 + 다양한 중증도 포함	Spectrum bias	**설계**
3	**표본 크기 산출**	DTA 공식 + 목표 Se/Sp + 최소 160 명 + 탈락률 반영	근거 불충분 판정	**설계**
4-1	**Index test**	정의 + 프로토콜 + 결과형식 + 품질기준 + Blind	Test review bias	**설계**
4-2	**Reference test**	임상진단 + 다학제팀(3 인) + 다중도구 + blind + 일치도 측정	Diagnostic review bias, Imperfect reference	**설계**
5	**결과 변수 선정**	1 차 outcome 변수 + 2 차 outcome 변수 선정	Selective reporting bias	**설계**
6	**평가 시점 및 운영**	구체적 시점 정의 + Index→Ref 순서 + T0·T1 포함	Temporal bias	**실행**
7	**평가자/진단팀 관리 및 눈가림**	전원 Ref 받음 + 양방향 blind + Index→Ref 순서 + Incorporation 차단	Verification/Diagnostic review bias	**실행**
8	**데이터 입력 및 품질관리**	표준화 CRF + 입력 가이드 + STARD 준수	Measurement bias	**실행**
9	**통계 분석 계획**	2×2 표 + Se/Sp(CI) + ROC/AUROC + Kappa + 민감도분석	Selective reporting bias	**분석**

핵심 차이점

- Domain 4 를 4-1(Index test), 4-2(Reference standard)로 분리
- Reference standard 신뢰도 측정(진단팀 간 일치도)이 필수
- 반복 측정(T0 + T1)으로 Reference 타당도 강화 및 Index test 예측력 평가
- 민감도 분석으로 Imperfect Reference Standard 문제 대응
- Domain 10-12 내용은 중재 연구와 동일함.

자가 점검 체크리스트

Domain 1. 연구 설계 유형

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	Prospective cohort study 설계를 채택하였다	☐	☐	☐
🔴	T0 시점과 T1 시점(6 개월 후)에 Index test 와 Reference standard 를 모두 수행하도록 계획하였다	☐	☐	☐
🔴	모든 연구대상자가 Index test 와 Reference standard 를 모두 받도록 설계하였다 (Complete verification)	☐	☐	☐
🔴	Index test 와 Reference standard 수행 순서를 사전에 명시하였다	☐	☐	☐
🔴	블라인드 유지 방법을 사전에 명시하였다	☐	☐	☐
🟡	3 개 이상의 기관이 참여하는 다기관 설계를 구성하였다	☐	☐	☐
🟢	독립적인 외부 데이터셋에서 Index test 성능을 재검증하는 외부 검증 계획을 수립하였다	☐	☐	☐
🟢	시간 간격을 두고 수집한 데이터에서 성능을 확인하는 시간적 검증 계획을 수립하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 5 개

Domain 2. 연구대상자 선정

● 기본 (필수), ● 권장 (상), ● 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
●	Target population 을 명확히 정의하였다	☐	☐	☐
●	포함 기준과 제외 기준을 명확히 정의하였다 (제외: 유전/대사/뇌병변 질환, Index test 수행 불가, 이미 확진받은 경우)	☐	☐	☐
●	경증, 중등도, 중증을 고르게 포함하여 Spectrum bias 를 예방하도록 계획하였다	☐	☐	☐
●	지역사회 기반 모집으로 경증 사례를 확보하도록 계획하였다 (지역사회 어린이집 연계 등)	☐	☐	☐
●	모집 및 선정 과정을 사전 정의하고 재현 가능한 형태로 문서화하였다	☐	☐	☐
●	STARD flow diagram 을 작성하는 계획을 수립하였다	☐	☐	☐

● **기본(필수) 항목 충족 현황**: 충족 ___ / 4 개

Domain 3. 표본 크기 산출

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	목표 민감도·특이도를 설정하였다 (예: Se ≥80%, Sp ≥80%)	☐	☐	☐
🔴	연구 표본에서 예상되는 발달장애 유병률을 고려하였다	☐	☐	☐
🔴	DTA 표본 크기 공식을 적용하였다: $n_cases = (Z^2\alpha/2 \times Se \times (1-Se)) / d^2$	☐	☐	☐
🔴	예상 탈락률(15-20%)을 반영하여 초기 모집 목표 수를 설정하였다	☐	☐	☐
🔴	SPACE 2 최소 기준 전체 160 명 이상(case/control 각 80 명)을 확보하도록 계획하였다	☐	☐	☐
🟡	층화변수별(연령, 중증도) 하위군 분석을 고려한 추가 표본 크기를 검토하였다	☐	☐	☐
🟢	외부 검증용 독립 데이터셋을 확보하는 계획을 수립하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 5 개

Domain 4-1. Index Test

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	Index test 를 명확히 정의하였다 (검사 방법, 장비, 소프트웨어 버전, Cut-off 값 명시)	☐	☐	☐
🔴	표준화된 프로토콜을 수립하였다 (검사 수행 조건 명시, Index test 후 Reference test 수행)	☐	☐	☐
🔴	이분형(양성/음성) + 연속형(확률값/점수) 결과를 모두 수집하여 ROC curve 를 작성하도록 계획하였다	☐	☐	☐
🔴	품질 관리 기준을 사전에 정의하였다 (부적절한 결과 제외 기준)	☐	☐	☐
🔴	Index test 판독자가 Reference standard 결과를 모르는 상태에서 판정하도록 블라인드를 계획하였다	☐	☐	☐
🟡	Index test 수행자의 자격 기준을 명시하였다	☐	☐	☐
🟡	사전 훈련 프로토콜을 마련하였다	☐	☐	☐
🟡	검사 일시, 장비 정보, 영상 품질 등급 등 메타데이터를 수집하도록 계획하였다	☐	☐	☐
🟢	영상 품질 평가자 간 일치도를 측정하는 계획을 수립하였다	☐	☐	☐
🟢	다양한 촬영 장비에서 성능을 평가하는 계획을 수립하였다	☐	☐	☐
🟢	영상 품질 등급별 민감도·특이도 변화를 분석하는 계획을 수립하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 5 개

Domain 4-2. Reference Standard

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	Reference Standard 를 명확히 정의하였다 (DSM-5 진단기준 적용)	☐	☐	☐
🔴	T1 시점(6 개월 후) 진단을 최종 Reference Standard 로 적용하도록 계획하였다	☐	☐	☐
🔴	3 인 이상 다학제 진단팀을 구성하였다 (소아정신과 전문의 1 인 + 임상발달심리사 1 인 포함, 발달장애 진단 경력 10 년 이상)	☐	☐	☐
🔴	다중 표준화 도구를 통합 사용하도록 계획하였다 (임상 진단 + ADOS-2 + ADI-R + 발달력 검토)	☐	☐	☐
🔴	다학제 진단팀이 Index test 결과를 모르는 상태에서 진단하도록 블라인드를 계획하였다	☐	☐	☐
🔴	다학제 진단팀 간 일치도를 측정 및 보고하도록 계획하였다 (2 팀 이상 운영 시 필수)	☐	☐	☐
🟡	ADOS-2 공인 평가자 자격을 보유한 전문의를 포함하였다	☐	☐	☐
🟢	Latent Class Analysis 를 적용하는 계획을 수립하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 6 개

Domain 5. 결과 변수 선정

● 기본 (필수), ● 권장 (상), ● 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
●	민감도, 특이도, AUROC 를 1 차 결과 변수로 지정하고 목표값을 설정하였다 (Se ≥80%, Sp ≥80%, AUROC ≥0.8)	☐	☐	☐
●	Index test-Reference standard 간 Kappa coefficient 를 산출하도록 계획하였다	☐	☐	☐
●	모든 지표에 95% CI 를 제시하도록 계획하였다	☐	☐	☐
●	2×2 분할표(TP, FP, TN, FN)를 작성하도록 계획하였다	☐	☐	☐
●	진단팀 간 일치도(Inter-rater reliability)를 측정하도록 계획하였다 (Fleiss' Kappa 또는 Cohen's Kappa, 95% CI 포함)	☐	☐	☐
●	반복측정 결과 변수를 정의하였다 (진단 안정성, 예측 민감도, 예측 특이도, test-retest reliability)	☐	☐	☐
●	하위군별 민감도·특이도를 분석하도록 계획하였다 (연령별, 중증도별, 질환 유형별)	☐	☐	☐
●	임상적 유용성 지표를 산출하도록 계획하였다 (Decision curve analysis, Calibration plot)	☐	☐	☐

● **기본(필수) 항목 충족 현황**: 충족 ___ / 5 개

Domain 6. 평가 시점 및 평가 운영

● 기본 (필수), ● 권장 (상), ● 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
●	모든 연구대상자와 기관에 동일하게 적용되는 구체적인 평가 시점을 사전에 정의하였다	☐	☐	☐
●	Index test 수행 후 Reference standard 를 수행하도록 순서를 명시하였다	☐	☐	☐
●	T0 시점 평가(Index test + Reference standard)와 T1 시점(6 개월 후) 평가를 계획에 포함하였다	☐	☐	☐
●	모든 연구대상자와 기관에 동일한 평가 간격과 assessment window 를 적용하도록 계획하였다	☐	☐	☐
●	표준화된 평가 운영 SOP 를 문서화하여 전 기관에 배포하도록 계획하였다	☐	☐	☐
●	평가 일정 이탈 발생 시 기록 및 보고를 의무화하는 절차를 수립하였다	☐	☐	☐
●	평가 시점 준수율을 모니터링하고 보고하는 체계를 마련하였다	☐	☐	☐

● **기본(필수) 항목 충족 현황**: 충족 ___ / 3 개

Domain 7. 평가자/진단팀 관리 및 눈가림

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	모든 대상자가 Index test 양성/음성 여부와 무관하게 동일한 Reference standard 를 받도록 설계하였다	☐	☐	☐
🔴	Index test 평가자에게 Reference standard 결과를 블라인드 처리하도록 계획하였다	☐	☐	☐
🔴	다학제 진단팀에게 Index test 결과를 블라인드 처리하도록 계획하였다	☐	☐	☐
🔴	Index test 수행 후 Reference standard 를 수행하는 순서를 확정하였다	☐	☐	☐
🔴	Index test 결과가 Reference standard 판정에 포함되지 않도록 차단하는 절차를 수립하였다	☐	☐	☐
🟡	Spectrum bias 예방을 위해 지역사회 기반 모집으로 경증 사례를 확보하도록 계획하였다	☐	☐	☐
🟢	블라인드 유지 성공 여부를 검증하는 Blinding Index 를 산출 및 보고하도록 계획하였다	☐	☐	☐
🟢	독립된 외부 진단팀을 활용하는 방안을 수립하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 5 개

Domain 8. 데이터 입력 및 품질 관리

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	전 기관이 동일하게 사용하는 표준화된 CRF 를 작성하였다	☐	☐	☐
🔴	변수 정의, 단위, 입력 형식, 결측값 처리 기준을 명문화한 데이터 입력 가이드라인을 작성하였다	☐	☐	☐
🔴	STARD 2015 체크리스트를 준수하도록 계획하였다	☐	☐	☐
🔴	STARD flow diagram 을 작성하도록 계획하였다	☐	☐	☐
🟡	범위 오류와 논리 불일치를 자동으로 검출하는 데이터 검증 절차를 적용하였다	☐	☐	☐
🟡	이중 입력 또는 주기적 Quality Control 을 수행하는 체계를 마련하였다	☐	☐	☐
🟡	기관별 결측값 발생 현황을 주기적으로 모니터링하고 보고하는 절차를 수립하였다	☐	☐	☐
🟢	개인정보 비식별화 프로토콜과 데이터 보존 계획을 수립하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 4 개

Domain 9. 통계 분석 계획

🔴 기본 (필수), 🟡 권장 (상), 🟢 이상 (최상)

구분	점검항목	충족	미충족	해당 없음
🔴	2×2 분할표, ROC curve 를 작성하도록 계획하였다	☐	☐	☐
🔴	민감도와 특이도를 95% CI 와 함께 산출하도록 계획하였다	☐	☐	☐
🔴	Index test-Reference standard 간 Kappa coefficient 를 산출하도록 계획하였다	☐	☐	☐
🔴	다학제 진단팀 간 Kappa coefficient 를 산출하도록 계획하였다	☐	☐	☐
🔴	민감도 분석을 계획하였다 (Reference standard 정의를 변경하여 Index test 성능 변화 확인)	☐	☐	☐
🟡	하위군별 민감도·특이도를 분석하고 Forest plot 으로 시각화하도록 계획하였다 (연령별, 중증도별, 질환 유형별)	☐	☐	☐
🟢	Calibration plot 을 작성하고 Hosmer-Lemeshow test 를 수행하도록 계획하였다	☐	☐	☐
🟢	Decision curve analysis 를 수행하여 임상적 의사결정에서의 Net benefit 을 평가하도록 계획하였다	☐	☐	☐

🔴 **기본(필수) 항목 충족 현황**: 충족 ___ / 6 개

부록 B. 중재 유형별 예시 모음

본 부록은 발달지연·발달장애 치료 연구를 진행하는 팀들이 각 치료 유형의 특성에 맞게 SPACE 2 가이드라인을 적용할 수 있도록 구체적인 예시를 제시하고 있으며, 각 치료 유형별로 설계에서 주의할 점, 자주 발생하는 오류, 권장하는 해결 방법을 Domain 별로 정리하였습니다.

□ 연구참여팀

No.	참여 연구팀	솔루션 명
1	**주식회사 솔리브벤처스**	ASD 아동 대상 ESDM 기반의 '스마트 교구'와 '모바일 앱'이 통합된 디지털 중재 플랫폼
2	**가톨릭대학교 산학협력단 ㈜디지털팜, 주식회사 셀바스에이아이**	발달장애 아동의 말소리 산출 능력 향상을 위한 프로젝트 **말해봄**
3	**푸르메재단 넥슨어린이재활병원**	**Switch On** (부제: 통합적 인지재활프로그램)
4	**동국대학교일산병원, ASK CST-K**	**WHO-Caregiver Skills Training** (CST)
5	**휴레이포지티브-세브란스병원-연세대학교 컨소시엄**	정밀 중증도 분류 기술을 활용한 ASD 조기선별 서비스 플랫폼 개발
6	**소중한언어랩 (언어발전소 x 소중한글 컨소시엄)**	양육자의 말 습관을 확장하는 LLM 기반 부모 참여형 솔루션 **“해빗톡(HabitTalk)”**
7	**뉴다이브 + 고려대학교 스펙트럼 컨소시엄**	자폐스펙트럼 장애 아동의 실행기능 증진을 위한 혼합형 디지털-아동 및 부모 중재 융합 모델
8	**주식회사 잼잼테라퓨틱스**	**잼잼 400: 핑크퐁과 잼잼 프랜즈** (실시간 모션인식 AI 기반, 뇌병변 및 자폐 아동을 위한 개인 맞춤형 소근육 재활치료 게임 솔루션)
9	**푸르메재단 넥슨어린이재활병원 정신건강의학과 정서행동치료팀**	**REDI: Ready for Education and Inclusion** “발달장애 아동의 건강한 첫 학교생활을 위한 맞춤 준비 프로그램”
10	**이른둥이 발달항해사**	**이른둥이**를 위한 가족중심 발달케어
11	**서울대학교병원 행동발달증진센터**	**아이 마음 D-탐사대:** 부모 매개 어플리케이션 기장 ABA 프로그램의 실증 연구
12	**H2H IN (Heart-to-Heart INteraction Lab)**	**H2H Remote Coaching:** 발달지연·발달장애 아동 언어발달을 위한 부모-아동 상호작용 기반 원격 부모코칭 솔루션

□ 부록 B 사용법:

- 자신의 중재 유형 섹션을 먼저 찾으세요
- Domain 별로 헷갈릴 수 있는 상황과 해결 방법을 확인하세요
- 실제 팀 번호가 명시된 예시를 참고하여 연구계획서를 작성하세요

※ **본문 연결:** 각 Domain 의 상세 원칙은 본문을 참조하세요. 이 부록은 원칙을 실제 중재에 어떻게 적용하는지에 초점을 맞춥니다.

□ 중재 유형별 핵심 주의 사항

중재 유형	핵심 주의사항	참여 연구팀 (No.)	권장 Domain
부모 매개	양육자 기준 별도 설정 가정 실습 모니터링 부모 관련 결과 변수 필수	4, 10, 11, 12	Domain 1, 2, 4, 5, 6, 7
디지털 중재	Active comparator 사용 자동 로깅 활용 평가자 블라인드 강화	1, 2, 6, 7, 8	Domain 1, 4, 6
언어·말소리 중재	접근법 구체적 명시 자발화 샘플 수집 일반화 평가	2, 6	Domain 4, 5
인지·실행기능 재활	인지 영역 명시 실험실 vs 일상 일반화 난이도 조절 알고리즘	3, 7	Domain 4, 5
학교준비 프로그램	학령 전환기 특정 장기 추적 평가 교사 협력	9	Domain 2, 5

※ 중재 유형 분류 안내

본 부록의 중재 유형 구분(부모 매개, 디지털 중재, 언어·말소리 중재, 인지·실행기능 재활, 학교준비 프로그램)은 연구 설계 참고를 위한 임의 분류입니다. 각 연구팀의 중재는 복합적 특성을 가질 수 있으므로, 본인 팀이 배정된 유형 외에도 중재의 다른 측면이 해당되는 유형이 있다면 해당 섹션을 함께 참고하시기 바랍니다.

1. 부모 매개 중재 (Parent-Mediated Intervention)

: 주 양육자가 치료자 역할을 수행하며, 전문가는 부모를 코칭하는 중재

Domain 1. 연구 설계 유형

Q1. 부모가 중재를 수행하는데, 부모를 블라인드 할 수 있나요?

- 부모는 블라인드 불가능함
- 평가자 블라인드는 반드시 유지해야 함
- 평가자가 해당 아동의 부모가 어떤 프로그램을 받았는지 모르는 상태에서 평가함

[원격 부모코칭 적용 예시]

설계:
- 중재군: 12 주 화상 부모코칭 + 가정 실습
- 대조군: 육아 정보 제공 (주 1 회 이메일)

블라인드:
- 부모: 블라인드 불가 (본인이 코칭 받으므로)
- 아동: 블라인드 불가 (영유아)
- 평가자: 블라인드 유지
  → 평가자는 "부모가 어떤 프로그램 받았는지" 모름
  → 평가 시 부모에게 프로그램 내용 질문 금지

Q2. 온라인 중재인데 대면 평가를 해야 하나요?

- 중재: 전국 어디서나 화상으로 제공
- 평가: 서울 병원에만 와야 한다면? → 지역 편향 발생

[Option1: 하이브리드 평가]

□ 필수 평가 (대면)
- Bayley-III (발달검사) - 표준화 필요
- ADOS-2 (자폐 평가) - 대면 관찰 필수

□ 선택 평가 (온라인)

- VABS-II (부모 보고)
- K-PSI (부모 스트레스)
- 화상으로 놀이 관찰 (15 분)

[Option2: 지역별 협력 기관]

(예시)

- 서울: 연세대 세브란스
- 부산: 부산대병원
- 대구: 대구파티마병원

⇒ 평가자 훈련 표준화 필수

⇒ 기관 간 신뢰도 측정 (Kappa ≥ 0.8)

▣ 본문 연결 → Domain 1 연구설계 유형: 블라인드 유지 방법

Domain 2. 연구대상자 선정

Q1. 아동 포함 기준은 정했는데, 부모 기준은 어떻게 설정하나요?

[아동 기준]

□ 포함

- 만 2-6 세 발달장애 또는 발달지연
- DSM-5 진단 또는 발달선별검사 양성

□ 제외

- 중증 뇌병변으로 WHO-CST 참여 불가
- 시청각 중복장애

[부모 기준]

□ 포함

- 주 양육자 1 인 이상 12 주 참여 가능
- 한국어 의사소통 가능
- 주 5 회 이상 가정 실습 수행 가능

□ 제외

- 심각한 정신질환 (조현병, 양극성장애) 치료 중
- 약물/알코올 중독 (최근 1 년)
- 가정폭력 이력 또는 아동학대 의심
- 인지기능 장애 (K-MMSE < 24)

Q2. 맞벌이 부부는 어떻게 하나요?

- 맞벌이 부부 → 낮 시간 중재 어려움
- 제외하면 → 대표성 문제

□ 중재 시간 유연화

- 주 양육자를 위한 저녁/주말 세션 추가
- 주말 세션 제공

※참고: 중재 프로토콜(중재 시간, 중재 방법 등)은 반드시 동일하게 진행해야 함.

▣ 본문 연결 → Domain 2 연구대상자 선정 및 무작위 배정: 포함/제외 기준 설정

Domain 4. 중재 프로토콜

Q1. 부모 교육은 어디까지 프로토콜에 포함되나요?

[X 불완전한 프로토콜 예시]

“12 주 동안 주 1 회 화상 부모코칭”

→ 프로토콜 내용이 너무 모호함

[O 올바른 프로토콜 예시]

[부모 교육 세션]

□ 구조

- 주 1 회, 60 분, 총 12 회
- 지난주 실습 피드백 (15 분)
- 이번주 목표 기술 이론 교육 (15 분)
- 실제 시연 및 역할극 (20 분)
- 이번주 가정 과제 안내 (10 분)

□ 주차별 주제

- Week 1-2: 아동 주도 놀이 기법
- Week 3-4: 의사소통 기회 만들기
- Week 5-6: 일상생활 속 언어 촉진

……

□ 가정 실습

- 주 5 회, 회당 20 분
- 부모-아동 놀이 상황
- 배운 기법 적용
- 영상 촬영 (주 1 회)

Q2. 부모마다 실력이 다른데, 충실도를 어떻게 측정하나요?

- 부모 A: 기법을 잘 사용함
- 부모 B: 잘 이해 못 함
- 둘 다 "중재군"인데 실제로는 다른 중재를 받은 셈

[충실도 측정 예시]

□ **1 단계: 구조적 충실도 (출석)**

- 측정: 화상 세션 참석 기록
- 목표: 12 회 중 10 회 이상 (≥83%)
- 판정: 자동

□ **2 단계: 가정 실습 빈도**

- 측정: 앱 로그 (일일 체크)
- 목표: 주 5 회 이상 (≥71%)
- 판정: 자동

□ **3 단계: 부모 기술 수행 질 (가장 중요함)**

- 측정: 영상 코딩
- 샘플: 전체 12 주 중 3 회 무작위 선정 (0 주, 6 주, 12 주)
- 코딩자: 독립 평가자 2 인
- 평가 항목:
  - ✓ 아동 주도 기회 제공 (회당 5 회 이상)
  - ✓ 아동 시도에 3 초 이내 반응 (≥80%)
  - ✓ 확장 기법 사용 (회당 3 회 이상)
  - ✓ 긍정적 강화 (회당 10 회 이상)
- 목표: 4 개 항목 중 3 개 이상 달성
- 평가자간 신뢰도: Kappa ≥0.7

▣ 본문 연결 → Domain 4 중재 프로토콜 표준화: 중재 충실도 측정

Domain 5. 결과 변수 선정

Q1. 아동 결과만 측정하면 안 되나요? 부모 결과도 필요한가요?

- 부모 매개 중재는 **부모 관련 결과 변수 포함 필수**

☐ **논리적 경로**

- 중재 → 부모 기술 향상 → 아동 발달 향상

☐ **만약 "부모 기술"을 측정 안 하면?**

- 효과가 없을 때 원인을 알 수 없음
  - ✓ "부모가 기술을 못 배웠나?"
  - ✓ "배웠는데 아동에게 효과가 없었나?"

[충실도 측정 예시]

☐ **1 차 결과 (아동)**

- VABS-Ⅱ Communication (사회적 의사소통)
- 목표 행동 빈도 (시선 맞춤, 요구하기)

☐ **1 차 결과 (부모) (필수)**

- MBRS (부모-아동 상호작용 질)
- PSOC (부모 양육 효능감)

☐ **2 차 결과 (부모)**

- PSI-SF (부모 양육 스트레스)
- PHQ-9 (우울)

Q2. 매개 분석은 어떻게 하나요?

- 독립변수: 중재군 vs 대조군
- 매개변수: 부모-아동 상호작용 질 (MRBS 점수)
- 종속변수: 아동 적응행동 (VABS-Ⅱ ABC)

□ 분석 방법

1. Baron & Kenny (1986) 단계
   - Step 1: 중재 → 아동 발달 (직접 효과)
   - Step 2: 중재 → 부모 기술 (a 경로)
   - Step 3: 부모 기술 → 아동 발달 (b 경로)
   - Step 4: 중재 + 부모 기술 → 아동 발달 (매개 효과)

2. PROCESS macro (Hayes, 2018)
   - Model 4 (Simple mediation)
   - Bootstrap 5000 회
   - 95% CI

▣ 본문 연결 → Domain 5 결과 변수 선정: 1 차 결과 변수, Domain 9 통계 분석 계획: 매개 분석

Domain 6. 평가 시점 및 간격

Q1. 부모 교육 중에도 아동이 계속 발달하는데, 언제 평가하나요?

□ **올바른 타이밍**

- T0: 0 주 (Baseline)
- T1: 12 주 (중재 직후) Primary endpoint
- T2: 24 주 (추적, Follow-up) → 효과 유지 확인

Q2. 부모 프로그램이 끝났는데 아동 효과는 나중에 나타나면?

□ **현상**

- Week 12: 부모 기술은 향상됨
- Week 12: 아동 변화는 아직 미미함
- Week 24: 아동 변화가 명확해짐 (부모가 계속 적용한 결과)

□ **해결책**

- Primary endpoint: T1 (12 주, 중재 직후)
  - ✓ 부모 기술 측정에 초점
- Secondary endpoint: T2 (24 주, 추적)
  - ✓ 아동 발달 변화 확인
  - ✓ 효과 유지 여부

□ **분석**

- T0 → T1: 즉각 효과
- T0 → T2: 지연 효과 및 유지
- T1 → T2: 효과 지속성

□ **문제**

- 고위험 영아는 월령에 따라 급격히 변화함
- 12 주 중재 동안에도 자연 발달이 발생함

□ **잘못된 타이밍**

- T0: 0 주 (중재 시작 전)
- T1: 6 주 (중재 중간) X 아직 중재 효과 나타나기 전
- T2: 12 주 (중재 종료)

▣ **본문 연결 → Domain 6 평가 시점 및 평가 운영: 평가 시점 설정**

Domain 7. 평가자/치료사 관리

Q1. 부모가 평가 때 “저희 코칭 프로그램 너무 좋았어요”라고 말하면?

[예방 프로토콜 예시]

□ 평가 전 - 보호자 교육

- 문자 발송 (평가 2 일 전):
  "평가 시 중재 프로그램 내용은 말씀하지 말아주세요.
  평가자가 객관적으로 평가해야 정확한 결과를 얻을 수 있습니다."
- 평가 당일 재안내:
  "오늘은 OO 의 현재 발달 상태만 평가합니다. 프로그램 만족도는 별도로 조사하겠습니다."

□ 평가 중 - 평가자 가이드

- 금지 질문:
  "집에서 어떤 앱을 사용하셨나요?"
  "프로그램이 어떠셨어요?"
- 허용 질문:
  "집에서 OO 와 어떻게 노시나요?"
  "요즘 OO 의 변화가 있나요?"

□ 평가 후 - Blinding check

- 평가자에게 질문:
  "이 아동이 어느 군인 것 같나요?" 질문
  → 정답률 50% (chance level)면 블라인드 성공

▣ 본문 연결 → Domain 7 평가자 관리 및 눈가림: 블라인드 유지 방법

---

2. 디지털 중재 (Digital Intervention)

: 앱, 웹 플랫폼, AI 알고리즘, 게임 기반 자동화된 중재를 제공하는 구조

Domain 1. 연구 설계 유형

Q1. 부모가 중재를 수행하는데, 부모를 블라인드 할 수 있나요?

[디지털 중재 설계 예시]

[비교군 설정]

- 중재군: 디지털 중재
- 대조군: 일반 교육 앱 (나이에 적절한 학습 앱)

→ 둘 다 “앱”이므로 위약 효과 통제

[평가자 블라인드 강화]

1. 평가실 환경

- 디지털 중재 관련 포스터 제거
- 디지털 중재 관련 교구 보이지 않게 보관

2. 보호자 사전 교육

- "평가 시 어떤 앱 사용했는지 말하지 마세요"

3. 평가자 질문 제한

- 금지 질문: "어떤 앱 쓰셨어요?"
- 대체 질문: "집에서 어떤 활동 하셨어요?"

4. 블라인드 확인

- 평가자에게 "이 아동이 어느 군인 것 같나요?" 질문

→ 정답률 50% (chance level)면 블라인드 성공

▣ **본문 연결 → Domain 1 연구 설계 유형: 평가 시점 설정, Domain 3 표본 크기 산출: 비교군 설정**

Q2. AI 알고리즘이 학습하면서 계속 바뀌는데, 연구 중에도 업데이트해야 하나요?

1. 연구 시작 시점 버전 고정

- LLM 모델명 및 버전 명시 / 앱 버전 고정
- 예: "OOO 모델 v2.1 (2025 년 1 월 배포 버전)"

2. 연구 기간 동안 고정 유지

- 새 버전 출시되어도 변경하지 않음
- 모든 연구대상자가 동일한 버전 사용

3. 예외 상황

- 보안 취약점 패치: 허용 (안전 문제)
- 성능 향상/기능 추가: 금지 (표준화 위배)

4. 연구계획서 명시 사항

"연구 기간(2025 년 1 월~6 월) 동안 앱 버전 및 알고리즘 고정.
보안 패치를 제외한 모든 업데이트 보류"

[알고리즘 고정 원칙]

▣ 본문 연결 → Domain 4 중재 프로토콜 표준화: 중재 정의 및 표준화

Domain 4. 중재 프로토콜

Q1. 앱 사용 데이터는 자동으로 쌓이는데, 이것만으로 충실도 측정이 되나요?

[예시 상황]

[아동 A]

- 앱 사용 시간: 20 분
- 세션 완료율: 100%
- 부모가 대신 터치함
- 아동은 TV 만 봄

[아동 B]

- 앱 사용 시간: 15 분
- 세션 완료율: 70%
- 아동이 집중해서 직접 함
- 부모가 적절히 코칭함

→ 로그만 보면 아동 A 가 더 좋아 보이지만 실제로는 B 가 올바른 중재임

[충실도 측정 예시]

☐ **1 단계: 자동 로그**

- 앱 접속 빈도: 주 5 회 이상
- 세션 완료율: ≥ 80%
- 평균 세션 시간: 20 분 이상

☐ **2 단계: 부모 일지 (매주)**

- 아동 참여도 (1~5 점 척도)
- 부모-아동 함께 한 시간 (분)
- 어려웠던 점

☐ **3 단계: 비디오 관찰 (샘플링)**

- 전체의 20% 무작위 선정 (주 1 회 제출)

- 평가 항목:
  - ✓ 아동이 직접 조작함 (vs 부모가 대신 조작함)
  - ✓ 아동이 화면 주시함 (vs 아동이 딴짓함)
  - ✓ 부모가 적절히 코칭함 (vs 부모가 과도하게 개입함)

▣ 본문 연결 → Domain 4 중재 프로토콜 표준화: 중재 충실도

Domain 5. 결과 변수 선정

Q1. 앱 내에서 수집한 성과 데이터를 결과 변수로 써도 되나요?

- 앱 내 과제 결과: 연습 효과인지, 실제 능력의 향상인지 구분 안 됨

□ **1 차 결과 (Proximal)**

- BRIEF-P (실행 기능 종합 평가)
- VABS-II Daily Living Skills (일상생활 기능)

□ **2 차 결과 (Secondary)**

- K-CBCL
- K-WISC-V 작업기억 지표 (만 6 세 이상)

□ **탐색적 분석**

- NEPSY-II 주의/실행기능 영역 (선택)
- 교사/부모 일상생활 평가

▣ **본문 연결 → Domain 5 결과 변수 선정: 1 차 결과 변수, 측정 도구 타당도**

Q2. 스크린타임이 늘어나는 것도 부작용으로 측정해야 하나요?

[예시]

□ **부모 질문:**

"우리 아이 하루 30 분만 화면 보라고 하는데, 연구 참여하면 20 분 더 보는 거 아닌가요?"

[대응 전략]

□ 동의서에 명시

1. 동의서 내용

"본 연구는 부모-아동 상호작용 기반 활동입니다. 아동이 혼자 화면을 보는 것이 아니라, 부모님과 함께 게임을 하는 구조입니다."

2. AAP 권장 사항

- 만 2-5 세: 고품질 프로그램 1 시간 이내
- 본 연구: 1 일 20 분 (권장 범위 내)
- 부모 동반 필수

□ 스크린타임 모니터링 (설문 조사 월 1 회)

1. 연구 앱 외 스크린 타임 (분/일)

- TV, 유튜브, 기타 앱

2. 총 스크린타임 계산

- 연구 앱 (20 분) + 기타

3. 안전 기준

- 총 스크린타임 > 90 분/일 → 경고
- 부모 상담 → 다른 스크린타임 줄이기 권장

▣ 본문 연결 → Domain 7 평가자 관리 및 눈가림: 안정성 모니터링

---

3. 언어, 말소리 중재

: 표현 언어, 수용 언어, 조음, 사회적 의사소통 능력 향상을 위한 중재

Domain 2. 연구대상자 선정

Q1. “언어 발달이 늦은 아동”이라고만 쓰면 안 되나요?

[말소리 산출 중재 포함 기준 예시]

[1 차 선별]

- 부모 보고: “말이 어눌하다” 또는 "발음이 부정확하다"

[2 차 확인 - 표준화 검사]

□ 포함

- U-TAP 조음 정확도 30-70%
- APAC 자음정확도 <-1.5 SD
- 연령: 만 3-6 세

□ 제외

- 조음 정확도 <30% **(너무 낮음 → 효과 측정 어려움 → 개별 치료 필요)**
- 조음 정확도 >70% **(거의 정상 → 천장 효과)**
- 구강 구조 이상 (구개열, 설소대 단축)
- 청각 장애 (순음청력검사 ≥25dB HL)
- 무발화 (표현 낱말 <10 개)

▣ 본문 연결 → Domain 2 연구대상자 선정 d 및 무작위배정: 포함/제외 기준

Q2. 자발화 샘플은 어떻게 수집하나요?

[자발화 샘플 수집 프로토콜 예시]

[준비]

□ **환경**

- 조용한 방, 책상과 의자 (아동용), 바닥 매트 (놀이용)

□ **장난감 세트 (통일)**

- 인형 가족 (4 개), 소꿉놀이 세트, 자동차 (3 대)
- 그림책 (2 권), 블록 (20 개)

[진행]

□ 시간: 30 분

□ 단계

1) Warm-up (5 분)
   - 자유 놀이
   - 평가자는 관찰만
2) 상호작용 (20 분)
   - 평가자가 아동과 함께 놀이
   - 개방형 질문 위주
   - "이거 뭐예요?" (X)
   - "여기서 뭐 하고 있어요?" (O)
3) 정리 (5 분)

[녹화]

□ 비디오 카메라 2 대

- 카메라 1: 아동 얼굴 정면
- 카메라 2: 전체 상황

□ 음성 녹음 (백업)

[전사 - Transcription]

□ 형식: CHAT (Codes for the Human Analysis of Transcripts)

□ 규칙: 모든 발화 전사, 불명료한 음소 표시, 제스처 기록, 맥락 정보 포함

[분석 지표]

1. MLU (Mean Length of Utterance)
   - 형태소 단위 평균 발화 길이

2. NDW (Number of Different Words)
   - 서로 다른 낱말 수

3. TNW (Total Number of Words)
   - 전체 낱말 수

4. 문법형태소 정확도
   - 조사 (-이/가, -을/를)
   - 어미 (-었어, -고 있어)

▣ 본문 연결 → Domain 5 결과 변수 선정: 평가 도구 선정

---

4. 인지, 실행기능 재활

: 주의집중력, 작업기억, 유연성 계획능력 등 인지능력 향상을 위한 중재

Domain 4. 중재 프로토콜

Q1. "인지훈련"이라고만 쓰면 안 되나요? 어떤 인지 영역인지 명시해야 하나요?

- 인지 영역 명시 - **필수**
- 구체적인 인지 영역 지정.

[인지 영역 명시 예시]

1. **주의집중력 (Attention)**
   - 지속 주의 (Sustained attention): 예) 숨은그림찾기 (3-5 분)
   - 선택 주의 (Selective attention): 예) 색깔만 보기 게임
   - 분할 주의 (Divided attention): 예) 걸으면서 공 주고받기

2. **작업기억 (Working Memory)**
   - 시각-공간 작업기억: 예) 카드 뒤집기 (Memory 게임)
   - 언어 작업기억: 예) 이야기 듣고 질문 답하기

3. **실행기능 (Executive Function)**
   - 억제 조절 (Inhibition): 예) 신호등 게임
   - 인지적 유연성 (Cognitive flexibility): 예) 규칙 바꾸기 게임 (처음엔 색깔로 분류하다가 나중엔 모양으로 분류하기)
   - 계획 수립 (Planning): 예) 샘플 보고 똑같이 블록 쌓기

4. **처리속도 (Processing Speed)**
   - 과제: 같은 것 빨리 찾기

5. **시공간 기능 (Visuospatial Function)**
   - 과제: 퍼즐 맞추기, 도형 회전하기, 위치 기억하기

▣ 본문 연결 → Domain 4 중재 프로토콜 표준화: 중재 정의

Q2. 아동마다 수준이 다른데 난이도를 어떻게 조절하나요?

[적응형 난이도 조절 예시]

[실시간 조정]

□ **규칙**

- 3 회 연속 정답 → 난이도 +1 단계
- 2 회 연속 오답 → 난이도 -1 단계

□ **난이도 레벨 (예: Go/No-Go task)**

- Level 1: 자극 제시 속도 2 초
- Level 2: 자극 제시 속도 1.5 초
- Level 3: 자극 제시 속도 1 초
- Level 4: 방해 자극 추가
- Level 5: 규칙 변경 (역전)

[세션별 시작점 조정]

□ Session 1: Level 1 부터 시작

□ Session 2: 전 세션 최종 레벨 -1 단계부터

□ Session 3: ...

[개인별 진행 기록]

□ 저장 데이터

- 세션별 도달 최고 레벨
- 각 레벨별 소요 시간
- 정답률 추이

→ 다음 세션 시작 난이도 자동 설정

▣ 본문 연결 → Domain 4 중재 프로토콜 표준화: 중재 용량 및 빈도

---

5. 학교준비 프로그램

: 유치원·초등학교 입학을 앞둔 발달장애 아동의 사회성, 학교적응 기술 향상을 위한 중재

Domain 2. 연구 대상자 선정

Q1. 이미 유치원 다니는 아동도 포함하나요?

[대상자 기준 예시]

□ 포함 기준

1. 6 개월~1 년 내 새로운 교육 기관 입학 예정

- 어린이집 → 유치원
- 유치원 → 초등학교
- 특수교육 → 통합학급

2. 현재 기관 적응에 어려움이 있는 경우

- 또래 상호작용 거의 없음
- 문제행동으로 개별 지도 필요
- 교사가 통합교육 어려움 호소

□ 제외 기준

1. 이미 통합학급에 완전히 적응 완료

- 또래 친구 3 명 이상
- 문제행동 거의 없음
- 교사가 추가 지원 불필요 판단

2. 중증도가 너무 높아 통합교육 불가능

- 공격 행동 주 5 회 이상
- 자해 행동 심각
- 의사소통 전혀 불가

▣ 본문 연결 → Domain 2 연구대상자 선정 및 무작위배정: 대상 집단 정의

부록 C. 평가영역별 표준화 검사도구

본 부록은 SPACE 2 참여 연구팀이 결과를 측정할 도구를 선택할 때 적절한 표준화 검사 도구를 선택할 수 있도록 평가 영역별로 국내외에서 신뢰도와 타당도가 확인된 도구 목록을 제시합니다. 각 도구의 적용 연령, 평가자, 소요 시간, 주요 하위척도, 국내 표준화 여부 등을 포함하였으며, 발달장애 아동 대상 연구에 적합성이 확인된 도구를 우선적으로 수록하였습니다.

□ 부록 C 사용법:

1. 도구 선택 기준

본 부록 수록 도구는 다음 기준을 충족합니다

- 국내 표준화 완료 또는 국제적 사용 도구
- 신뢰도 계수(Cronbach's α) 0.80 이상
- 타당도 검증 완료
- 발달장애 아동 적용 가능

2. 평가 영역 구성

- 적응행동 및 일상생활 기술
- 언어·의사소통
- 인지·실행기능
- 사회성·정서
- 문제행동
- 감각·운동
- 부모·가족 관련 변수

3. 도구 선택 절차

- **Step 1:** 연구의 1 차 결과 변수에 해당하는 평가 영역 확인
- **Step 2:** 대상 아동 연령대와 도구의 적용 연령 매칭
- **Step 3:** 평가자 자격(전문가/부모/교사) 확인
- **Step 4:** 다기관 연구 시 모든 기관 동일 도구 버전 사용

전체 평가영역 통합표

대분류	세분류 핵심	주요 평가도구
언어·의사소통	수용언어, 표현언어, 어휘, 문장, 이야기, 발성	▪ Receptive and Expressive Vocabulary Test, REVT ▪ Sequenced Language Scale for Infants, SELSI ▪ Preschool Receptive-Expressive Language Scale, PRES ▪ Korean Narrative Assessment, KONA ▪ Mean Length of Utterance in Words, MLUw ▪ Number of Different Words, NDW ▪ Length of Longest Utterance in Words, LLUw ▪ Praat Speech Analysis Software; Computerized Language Analysis, CLAN
말소리·조음	자음정확도, 조음오류, 말운동, 발성시간	▪ Assessment of Phonology and Articulation for Children, APAC ▪ Urimal Test of Articulation and Phonology, U-TAP ▪ Speech Mechanism Screening Test for Children, SMST-C ▪ Maximum Phonation Time, MPT; Diadochokinetic Rate, DDK
인지·학습	전반 인지, 시지각, 수개념, 학습준비	▪ Kaufman Assessment Battery for Children, Second Edition, KABC-II ▪ Bayley Scales of Infant and Toddler Development, Third Edition, Bayley-III ▪ Beery-Buktenica Developmental Test of Visual-Motor Integration, Sixth Edition, Beery VMI-6 ▪ Korean Developmental Test of Visual Perception, Third Edition, K-DTVP-3

		▪ Developmental Assessment of Readiness for Basic Academic Skills, DARBAS ▪ Basic Academic Skills Assessment, BASA
실행기능	억제, 작업기억, 유연성, 계획, 주의	▪ Executive Function Touch, EF Touch ▪ Head-Toes-Knees-Shoulders Task, HTKS ▪ Stroop Color and Word Test; Attention Control Evaluation, ACE ▪ Working Memory Test, WMT ▪ Children's Trail Making Test, C-TMT ▪ Continuous Performance Test, CPT ▪ Behavior Rating Inventory of Executive Function, BRIEF ▪ Behavior Rating Inventory of Executive Function, Second Edition, BRIEF-2
사회성	사회적 반응성, 공동주의, 또래상호작용	▪ Social Responsiveness Scale, Second Edition, SRS-2 ▪ Social Responsiveness Scale, SRS ▪ Autism Diagnostic Observation Schedule, Second Edition, ADOS-2 ▪ Childhood Autism Rating Scale, Second Edition, CARS-2 ▪ Brief Observation of Social Communication Change, BOSCC ▪ Penn Interactive Peer Play Scale, PIPPS
적응행동	자조, 일상생활기술, 사회화, 학교 적응	▪ Vineland Adaptive Behavior Scales, Third Edition, Vineland-3 ▪ Korean Vineland Adaptive Behavior Scales, Third Edition, K-Vineland-III ▪ Vineland Adaptive Behavior Scales, Third Edition, VABS-3 ▪ Korean Vineland Adaptive Behavior Scales, Second Edition, K-VABS-II ▪ Pediatric Evaluation of Disability Inventory, PEDI ▪ Pediatric Evaluation of Disability Inventory-Computer Adaptive Test, PEDI-CAT

대근육·소근육	영아 운동, 상지기능, 손기능, 악력, 집기	▪ General Movements Optimality Score-Revised, GMOS-R ▪ Motor Optimality Score-Revised, MOS-R ▪ Hammersmith Infant Neurological Examination, HINE ▪ Gross Motor Function Measure, GMFM ▪ Melbourne Assessment 2, MA2 ▪ Box and Block Test, BBT ▪ Tip Pinch Strength Test; Grip Power Test
정서·행동	문제행동, 도전행동, 정서조절, 행동기능	▪ Child Behavior Checklist, CBCL ▪ Korean Child Behavior Checklist, K-CBCL ▪ Aberrant Behavior Checklist, ABC ▪ Korean version of the Challenging Behavior Scale for Children, K-CSCB ▪ Questions About Behavioral Function, QABF ▪ Childhood Challenging Behavior Scale, Second Edition, CCBS-2 ▪ Emotion Dysregulation Inventory, EDI ▪ Korean ADHD Rating Scale, Fifth Edition, K-ARS-5
부모 역량·심리	양육지식, 효능감, 스트레스, 우울, 애착	▪ Caregiver Knowledge and Skills Test, CKST ▪ Parent Knowledge and Skills Test, PKST ▪ Korean Parenting Stress Index, Fourth Edition, K-PSI-4 ▪ Korean Parenting Stress Index, Fourth Edition-Short Form, K-PSI-4-SF ▪ Korean Parenting Sense of Competence Scale, K-PSOC ▪ Center for Epidemiologic Studies Depression Scale, CES-D ▪ Beck Depression Inventory, Second Edition, BDI-2 ▪ Korean Maternal Postnatal Attachment Scale, MPAS-K

부모-아동 상호작용	부모 반응성, 언어자극, 공동주의, 차례 주고 받기	▪ Adult-Child Interaction Scale, ACI ▪ Evaluation of Social Interaction, ESI ▪ EUDICO Linguistic Annotator, ELAN ▪ Computerized Language Analysis, CLAN ▪ Praat Speech Analysis Software
사용성·순응도	앱 사용성, 접속, 미션, 만족도, 수용성	▪ user version of the Mobile Application Rating Scale, uMARS ▪ Korean Quebec User Evaluation of Satisfaction with Assistive Technology, Version 2.1, K-QUEST 2.1 ▪ app usage logs ▪ mission completion rate ▪ session attendance rate ▪ satisfaction survey
안전성·실행가능성	이상반응, 수행률, 중도탈락, 영상확보	▪ 이상반응률 ▪ SpO_2 ▪ 심박수 ▪ 수행 세션 비율 ▪ 영상 확보율
생체·AI 지표	안저, EEG, MRI, 움직임 AI, 모델성능	▪ Area Under the Receiver Operating Characteristic Curve, AUROC ▪ 민감도 ▪ 특이도 ▪ Electroencephalography, EEG ▪ functional Magnetic Resonance Imaging, fMRI ▪ Diffusion Tensor Imaging, DTI ▪ pose estimation

가족기능·삶의 질	가족기능, 아동 삶의 질, 부모 부담	▪ Korean Family Functioning Measure, K-FFM ▪ KIDSCREEN-52 Health-Related Quality of Life Questionnaire ▪ TNO-AZL Preschool Children Quality of Life Questionnaire, TAPQOL ▪ Korean Parenting Stress Index, K-PSI