

「兒童健康手冊」發展監測項目之信效度研究

廖華芳¹ 鄭玲宜¹ 謝武勳² 楊銘欽³ 鄒國蘇⁴ 蔡昆瀛⁵

本研究之目的為探討國民健康局編印之「兒童健康手冊」內發展狀況題目(Developmental Surveillance Items of Child Health Pamphlet, DICHP)之信效度。信度研究設計採方法學研究，效度採病例對照研究。以方便取樣方式，選取1歲至3歲3個年齡層，每個年齡層一般兒童與遲緩兒童人數比例為1-3:1，共395名兒童參與效度研究，其中194名兒童於一週內測試2次，以分析信度。經審視DICHP題目，將2題語意不清楚者略作文字修改，編製DICHP初修版，每個年齡層仍維持6題。請兒童父母或主要照顧者填寫DICHP初修版，據以分析內部一致性與各題目及總分之再測信度、接受者操作特徵(receiver operating characteristic, ROC)曲線下面積、最佳切截點之敏感度與特異度、多層次概率比、已知族群效度(known-groups validity)。結果顯示DICHP初修版內部一致性($\alpha=0.74-0.86$)與再測信度(斯皮爾曼相關係數 $=0.72-1.00$; $p < 0.001$)可；由ROC曲線得其篩選正確機率 $=0.82-0.86$ ($p < 0.001$)；3個年齡層之敏感度範圍為0.69-0.80，特異度為0.76-0.89；綜合各年齡層之多層次概率比，5-6分者(0-1題不通過者)，其概率比皆小於1； $\leq 2-4$ 分者(2-4題以上不通過者)其概率比皆為無限大，符合可幫助確定陽性診斷之標準；遲緩兒童之分數顯著低於一般兒童，具已知族群效度。所以，3個年齡層DICHP之信效度在可接受範圍，可當多階段篩選之第一層工具。根據多層次概率比，建議 $\geq 2-4$ 題不通過，應轉介診斷測試；1-3題不通過進一步二篩；5-6題通過視為正常。

關鍵字：發展遲緩，篩選測試，信效度，概率比，兒童
(台灣醫學 Formosan J Med 2008;12:502-12)

前 言

在兒童發展評估過程中，發展評量工具尤其重要[1]。「發展篩檢」是早期發現之重要管道。美國小兒科醫學會提出發展篩檢的工作準則中，使用良好的篩檢工具是執行篩檢工作的必要利器[1]，此外，良好篩檢工具必須符合簡單、快速、容易解釋(通不通過、是否轉介的標準明確便於讓非專業人員執行)、良好信度、良好效度、能定期篩檢等標準[2]。2003年修正公布的「兒童及少年福利法」與2004年6月修訂公布的「特殊教育法」皆規定發展遲緩兒童之早期療育，應由醫療主管機關召

集，並結合醫療、教育、社政等主管機關，共同規劃及辦理早期療育工作。2007年7月修訂公布之「身心障礙者權益保障法」第十八條規定衛生主管機關負責彙送疑似身心障礙者、發展遲緩或異常兒童資訊予通報轉介中心。為達到早期發現、早期通報之目的，早期發展評量(early development evaluation)的規劃與執行，就變成推動早期療育的首要工作。

過去研究顯示，台灣地區「兒童健康手冊」之利用情形不佳，衛生單位應檢討此手冊之實用性及可行性[3]；而行政院衛生署國民健康局也於2005年1月新編「兒童健康手冊」；且配合全民健康保


台大醫學院物理治療學系暨研究所、台大醫院復健醫學部；¹台北榮民總醫院復健醫學部；²台灣大學醫學院小兒部新生兒科；³台灣大學公共衛生學院醫療機構管理研究所；⁴台北市立聯合醫院婦幼院區兒童精神科；⁵台北市立教育大學特殊教育學系

受文日期 民國96年10月30日

接受刊載 民國97年2月26日

通訊作者聯絡處：廖華芳，台大醫學院物理治療學系暨研究所，台北市徐州路17號3樓

表一：3 個年齡層「兒童健康手冊」之發展題目初修版(信效度資料收集用)

年齡層	題目	發展領域
12 個月 (11.5-17.5)	1. 能不扶任何東西，自己行走了嗎?	粗動作
	2. 會用拇指和食指的指尖捏起小東西了嗎?(例如葡萄乾、豆子)	精細動作
		
	3. 能瞭解幾個單字的意義嗎?例如問他「狗狗呢?」「姊姊呢?」，會轉頭找尋標的物或人	語言
	4. 會說一、二個有意義的單字嗎?例如「抱抱」「媽媽」	語言
	5. 會用姿勢表達自己的需要嗎?例如用手指或拉扯大人的衣服	語言
18 個月 (17.5-23.5)	6. 將玩具當著寶寶的面藏在手帕或床單下，他會掀開找出玩具嗎?	認知
	1. 能由大人牽著一隻手或自己扶著欄杆走上樓梯嗎?	粗動作
	2. 會自己爬進大人的椅子坐好嗎?	粗動作
	3. 會說五個以上有意義的單字了嗎?	語言
	4. 會指出至少三個身體部位了嗎?	認知 ¹
	5. 會聽從簡單的口頭指令嗎?例如「拿去給哥哥」「去拿鞋鞋」	語言
24 個月 (23.5-35.5)	6. 會模仿大人使用家裡的用具或做家事嗎?例如聽電話、拿梳子梳頭、拿布擦桌子?	社會/認知
	1. 能由大人牽著一隻手或自己扶著欄杆走下樓梯嗎?	粗動作
	2. 會雙腳離地跳躍嗎?	粗動作
	3. 會將幾塊積木堆高嗎?	精細動作 ¹
	4. 會將兩個單字組成短句嗎?例如「媽媽抱」「看狗狗」	語言
	5. 會正確指認一、二樣圖片中的東西或動物嗎?例如球、狗	認知
	6. 會自己脫掉鞋襪嗎?	生活自理

¹ 修改之題目，後有子題目未列在本表格中

原題目：18 個月-4. 會指出至少三個身體部位了嗎?例如眼睛、嘴巴、手
24 個月-3. 會將幾塊積木或類似的方塊物體堆高嗎?

險預防保健服務之給付時程，在 9 個年齡層(兒童 1、2~3、6~7、9~10 個月，與 1~1.5、1.5~2、2~3、3~4、4~7 歲)皆有 2-6 個發展狀況題目 (Developmental Surveillance Items of Child Health Pamphlet, DICHP)，由家長勾選，再交由醫師判讀。然如何判讀這些發展狀況，並無適當指引。由於發展遲緩越早發現越好，因此本研究希望針對 3 歲以下兒童進行；但臨床上一歲之前之發展遲緩兒童人數有限，乃以 1-3 歲兒童為研究族群。本研究之目的為探討 1-3 歲兒童 DICHP 之信效度，包括：(1)內部一致性與各發展項目及各年齡層總分之再測信度；(2)各年齡層總分之效度，即已知族群效度(known group validity)、接受者操作特徵(receiver operating characteristic, ROC)曲線下面積(area under curve, AUC)、多層次概率比 (Multilevel

likelihood ratio)；(3) 根據 ROC 與多層次概率比分析結果提出臨床使用建議。

研究對象與方法

一. 研究設計

信度採方法學研究設計 (methodology research)，效度以病例對照研究(case control study) 方法。吳等人依過去美國相關研究結果，推估發展遲緩兒童盛行率約為 4~9%[4]，因此若以連續性樣本，在低盛行率下要收集大量兒童資料才有足夠之發展遲緩兒童以分析 DICHP 之效度；本研究乃以病例對照研究，希望在短期能得到 DICHP 之效度資料，以為臨床運用之參考。

二. 研究對象

共分三個年齡組，計 12 個月組(年齡 11 個月

16 天-17 個月 15 天)、18 個月組(17 個月 16 天-23 個月 15 天)、與 24 個月組(23 個月 16 天-35 個月 15 天)。在 2 歲以下早產兒童,以矯正月齡分組。以方便取樣方式,每一個年齡層以一般兒童約為遲緩兒童之 1-3 倍比例選取。收取的條件,一般兒童為在收集資料時未被診斷有發展遲緩或疑似發展遲緩;遲緩兒童為被診斷為發展遲緩、疑似發展遲緩、或發展障礙,已接受早期介入或即將接受早期介入。發展遲緩之診斷為門診醫師或兒童發展評估專業團隊根據觀察與家長晤談、身體檢查與/或標準化發展診斷量表而確認。

收取個案方式,一般兒童之收取為研究人員於台大醫院小兒部小兒健康門診,台北榮總復健部遲緩兒童聯合評估中心,臺北市聯合醫院婦幼院區,嘉義莊凱全小兒科診所,初步篩選與收案條件符合,並徵求門診等候個案之父母填寫發展問卷,或寄問卷給有符合收案條件之研究團隊之親戚朋友。遲緩兒童之來源為台大醫院小兒部、復健部,台北榮總復健部,台北聯合醫院婦幼院區,臺北市早期療育綜合服務中心,台北市新光醫院復健部,台北市社會局及嘉義莊凱全小兒科診所,在這些單位中接受早期介入或即將接受早期介入,並願意參與此研究者。本研究計畫已經台大醫院研究倫理委員會審查通過,且所有參與兒童之父母皆填寫同意書。

三. 流程

在進行信效度研究之先,先由研究者審視 3 個年齡層 DICHHP 之各題目,經由研究團隊討論,並略加修改,編製 DICHHP 初修版(見表一)。除原有之題目,另就題目敘述不清楚者,再加子題,以作為將來修改之參考;分析信效度時以每個年齡層原 6 題之修正題目分析。效度方面,請一般兒童與發展遲緩兒童之父母或主要照顧者填寫 DICHHP 初修版與基本資料。信度方面,一個星期前後填寫兩次以得再測信度。

四. 資料分析

兒童之行為若符合 DICHHP 之題目描述給與 1 分,若不符合則得 0 分,因此每個兒童 DICHHP 總分之範圍為 0-6 分,分數越高發展越好。信度方面,除以傳統之內部一致性(Cronbach α 係數)分析外,尚對個別項目計算各年齡層兒童之一致性

(agreement)與卡帕係數(Kappa)。一致性為兩次測試結果一致者佔總測試人數之百分比,百分比越高則再測信度越高。卡帕係數之解釋為:極低一致性(卡帕係數範圍 0.00-0.20),一致性尚可(fair)(0.21-0.40),中度一致性(0.41-0.60),高度一致性(substantial)(0.61-0.80),近完美一致性(0.81-1.0)[5]。各年齡層 DICHHP 總分以斯皮爾曼相關係數檢定其再測信度。效度方面,進行以下分析:(1)篩選正確機率(screening accuracy);(2)多層次概率比;(3)已知族群效度。 α 值設為 0.05。

篩選正確機率以 AUC 推估[6]。由 ROC 曲線分析可得測驗之最佳切截點,再由列聯表得此最佳切截點之敏感度(sensitivity)與特異度(specificity)[6]。ROC 曲線是以「錯誤接受率」(false positive fraction,即 1-特異度)為橫座標,「真陽率」(true positive fraction,即敏感度)為縱座標,由無數切截點求出的無數對(1-特異度)和敏感度作圖構成曲線。AUC 愈大,測驗就愈有區辨效果,代表測量準確率愈高;當在某個切截點,其敏感度和特異度均為 1 時,標記點會在曲線的左上角,AUC 等於 1,ROC 曲線最完美[6]。因此 ROC 曲線上最接近左上角之點,即為最佳切截點。由最佳切截點以列聯表可得敏感度、特異度。再根據(敏感度)/(1-特異度)公式算出陽性概率比(positive likelihood ratio, LR+)及(1-敏感度)/(特異度)公式算出陰性概率比(negative likelihood ratio, LR-)[7]。

由於不同的切截點通常會使敏感度與特異度互有增減,因此在綜合考量設定切截點時,還會檢視「概率」(likelihood)[7]。由於多截點之參考價值高於單截點[7],因此本研究嘗試使用多層次概率比來分析 DICHHP 之篩選效度。概率比(likelihood ratio, LR),為真正發展遲緩兒童中得該測試值之機率除以正常兒童中得該測試值之機率。一般來說,測試後勝算(post-test odds)為測試前勝算(pre-test odds)乘以 LR 值。測試前勝算為測試前機率除以(1-測試前機率),測試後勝算則為測試後機率除以(1-測試後機率)。因此 LR 值大小,表示該測試結果可以改變測試前後機率之程度[8]。LR 值等於 1,表示測試前後機率並無改變;LR 值大於 1,表示測試後機率會增加,值越大其增加程度越高,即診斷影響力(diagnostic impact)高;相反的,

表二：參與效度研究之一般兒童與遲緩兒童之人口學基本資料

	遲緩兒童 (n=181)	一般兒童 (n=214)
平均年齡(±SD)(月)	23.9±6.8	20.9±7.1
性別 男：女(人數)	112:69	94:120
妊娠週數 ≤31 週：32-36 週：≥37 週(人數)	21:35:116	18:35:158
出生體重(人數)		
≤1500 公克：1501-2499 公克：≥2500 公克	19:32:122	16:22:171
母親年齡(±SD)(歲)	33.4±4.4	33.1±4.2
母親職業 職業婦女：非職業婦女(人數)	79:96	146:67
母親國籍 台籍：非台籍(人數)	168:12	201:13
母親教育程度 ≤國中：高中：≥大專(人數)	9:52:116	6:42:164
父親年齡(±SD)(歲)	35.8±5.1	35.8±5.0
父親職業 軍公教：自由服務：工：其他(人數)	15:69:34:58	44:82:28:55
父親國籍 台籍：非台籍(人數)	178:1	207:6
父親教育程度 ≤國中：高中：≥大專(人數)	8:49:118	4:39:168
家庭結構 大家庭：折衷家庭：小家庭(人數)	35:41:104	26:67:119
地理分布 北：中：南(人數)	153:1:2	119:11:53
收案來源 醫學中心：非醫學中心(人數)	123:56	124:89

註：部份變項資料有缺失，因此總數並非與收案人數同

LR 值小於 1，表示其測試後機率會減少，值越小其減少程度越明顯，因此也有較高之診斷影響力。因此 LR 值大於 10 或小於 0.1 表示改變測量前後機率之程度高，診斷影響力屬高度(large)，可幫助確定陽性診斷(SpPin)或確定陰性診斷(SnNout)；5-10 或 0.1-0.2 為中度(moderate)，2-5 或 0.2-0.5 為輕度有時重要(small, sometimes important)，LR 值 1-2 或 0.5-1 之測試分數則診斷影響力輕度重要性少(small, rarely important)[8]。

已知族群效度為對已知不同之族群進行測試，比較兩族群之測試值，以了解此工具對此不同族群之區別程度[9]。由於本研究一般兒童與遲緩兒童之 DICHP 分數皆非常態分佈，乃以 Mann-Whitney 檢定一般兒童與發展遲緩兒童之 DICHP 總分有無顯著差別。

研究結果

一. 參與研究者之基本資料

共 395 名兒童參與效度研究(214 一般兒童與 181 名遲緩兒童)，其中 109 名一般兒童與 85 名遲緩兒童參與信度研究。這些兒童及其家庭背景之基

本資料如表二所示。一般兒童填答者以母親居多(85%)；母親平均年齡約 33 歲，台灣籍(94%)、教育程度大專(含)以上(77%)居多；父親平均年齡約 36 歲，台灣籍(97%)、教育程度大專(含)以上(80%)居多。遲緩兒童填答者也以母親居多(82%)；母親平均年齡約 33 歲，台灣籍(93%)、教育程度大專(含)以上(66%)居多；父親平均年齡約 36 歲，台灣籍(99%)、教育程度大專(含)以上(67%)居多。

在兒童之低妊娠週數人數比例($\chi^2_{(df=2)} = 2.7; p=0.26$)、母親之年齡 ($t = 0.65; p=0.52$)、國籍($\chi^2_{(df=1)} = 0.06; p=0.81$)、與教育程度($\chi^2_{(df=5)} = 6.4; p = 0.27$)方面，兩組兒童並無顯著差別。在兒童之年齡($t = 4.3; p < 0.001$)、男童比例($\chi^2_{(df=1)} = 12.8; p < 0.001$)、與低出生體重兒人數比例($\chi^2_{(df=2)} = 7.0; p < 0.05$)方面，則遲緩兒童顯著高於一般兒童；遲緩兒童其母親為職業婦女之比例也低於一般兒童($\chi^2_{(df=2)} = 21.6; p < 0.001$)。

為了解本研究之一般兒童與其母親其人口學特性是否與台灣地區之同年齡層兒童與其母親有無顯著差別，以內政部戶政司網頁(網址 <http://www.ris.gov.tw/docs/f4a.html>)之資料當母群體，根據卡方適合度檢驗 (Chi-Square

表三：「兒童健康手冊」發展題目初修版個別題目於遲緩兒童與一般兒童之再測信度

年齡層	題目	遲緩兒童			一般兒童		
		一致性(%)	卡帕係數	p 值	一致性(%)	卡帕係數	p 值
12 個月 (遲緩兒童 16 位) (一般兒童 45 位)	1	94	0.82	.001	96	0.89	<.001
	2	94	0.86	.000	100	1.00	<.001
	3	81	0.61	.013	98	0.66	<.001
	4	87	0.71	.004	91	0.66	<.001
	5	94	0.85	.001	100	1.00	<.001
	6	87	0.58	.024	100	1.00	<.001
18 個月 (遲緩兒童 26 位) (一般兒童 27 位)	1	100	1.00	<.001	100	-	-
	2	100	1.00	<.001	96	-	-
	3	92	0.83	<.001	89	0.60	.002
	4	92	0.80	<.001	100	1.00	<.001
	5	96	0.91	<.001	100	-	-
	6	96	0.90	<.001	100	-	-
24 個月 (遲緩兒童 43 位) (一般兒童 37 位)	1	100	1.00	<.001	100	-	-
	2	98	0.95	<.001	100	-	-
	3	95	0.83	<.001	100	-	-
	4	88	0.75	<.001	100	-	-
	5	98	0.94	<.001	100	-	-
	6	100	1.00	<.001	100	1.00	<.001

表四：「兒童健康手冊」發展題目初修版各年齡層總分之再測信度

	遲緩兒童			一般兒童		
	人數	r _s ¹	p 值	人數	r _s ¹	p 值
12 個月	14	0.88	<.001	45	0.86	<.001
18 個月	23	0.99	<.001	26	0.72	<.001
24 個月	37	0.94	<.001	35	1.00	-

¹ 斯皮爾曼相關係數(Spearman correlation coefficient)

goodness-of-fit test), 結果顯示在兒童性別上與母群體無顯著差異($\chi^2_{(df=1)}=1.11; p=0.29$), 然本研究一般兒童樣本之母親台籍比例較母群體高($\chi^2_{(df=1)}=8.6; p < 0.01$), 母親教育程度也較母群體高($\chi^2_{(df=1)}=120.6; p < 0.001$)。

遲緩兒童之診斷包括：全面性發展遲緩、語言發展遲緩、動作發展遲緩、腦性麻痺、自閉症、注意力不足過動症、癲癇、染色體異常等；其中以全面性發展遲緩居多(61%)，其次為語言發展遲緩(11%)與腦性麻痺(7%)。

二. 信度

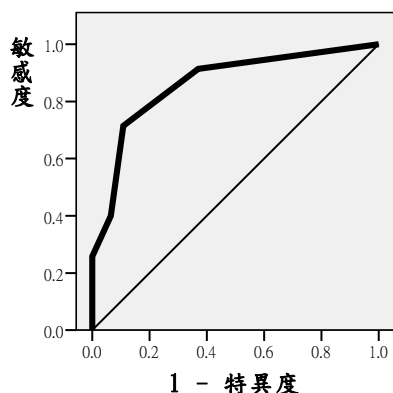
於內部一致性方面，於 12、18 與 24 個月三個年齡層之 Cronbach α 係數分別為 0.74、0.86 與 0.84，為可接受範圍。於再測信度方面，一般兒童

與遲緩兒童各年齡層各題目之再測信度如表三所示。一般兒童各題目之一致性介於 89%-100%；卡帕係數範圍 0.60-1.00，兩次測試間有高度至接近完美相關(p 值<0.05)。遲緩兒童之一致性介於 81%-100%；卡帕係數範圍 0.58-1.00；兩次測試間有中度至接近完美相關(p 值<0.05)。

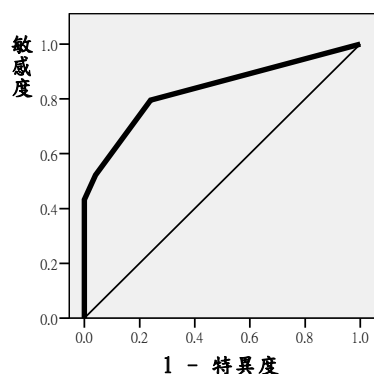
兩組兒童各年齡層總分之再測信度如表四所示，一般兒童斯皮爾曼係數範圍 0.72~1.0(p 值<0.001)，遲緩兒童斯皮爾曼係數範圍 0.88-0.99(p 值<0.001)，兩次測試間皆有顯著相關。

三. 篩選正確機率

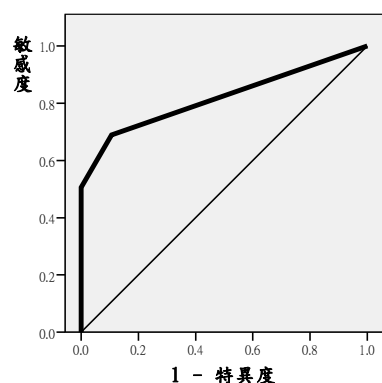
圖一為 DICHP 初修版於 3 個年齡層之 ROC 曲線。12 個月年齡層的兒童其 AUC 為 0.86(95% CI, 0.79-0.94; $p < 0.001$)，最佳切截點為 5，即小於 5 即



圖一 a: 12 個月 AUC=0.86; 95%信賴區間, 0.79-0.94; p 值 < 0.001, 最佳切截點為 5



圖一 b: 18 個月 AUC=0.83; 95%信賴區間, 0.75-0.92; p 值 < 0.001, 最佳切截點為 6



圖一(c): 24 個月 AUC=0.82; 95%信賴區間, 0.75-0.89; p 值 < 0.001, 最佳切截點為 6

圖一：三個年齡層兒童健康手冊發展題目初修版之 ROC 曲線

是懷疑有發展遲緩，在此切截點下其敏感度為 0.71，特異度為 0.89，LR+ 為 6.4，LR- 為 0.32；18 個月之 AUC 為 0.83(95% CI, 0.75-0.92; p 值 < 0.001)，最佳切截點為 6，在此切截點下其敏感度為 0.80，特異度為 0.76，LR+ 為 3.3，LR- 為 0.26；24 個月之 AUC 為 0.82(95% CI, 0.75-0.89; p 值 < 0.001)，最佳切截點為 6，在此切截點下其敏感度為 0.69，特異度為 0.89，LR+ 為 6.3，LR- 為 0.35。因此綜合 3 個年齡層之篩選正確機率為 0.82-0.86。

四. 多層次概率比

表五為 DICHP 初修版於 3 個年齡層之多層次概率比。隨著 DICHP 初修版分數之增加，其 LR 有逐漸減少之傾向，但於 12 個月年齡層可能因遲緩個案數不足，於 3-4 分間有一些不一致傾向。12 個月年齡層在總分小於等於 2 分者 LR 都是無限大，3-4 分者(2-3 題不通過)LR 大於 1，5-6 分者 LR 小於 1；18 個月年齡層小於等於 3 分者 LR 都是無限大，4-5 分者(1-2 題不通過)LR 大於 1，6 分者(全通過)LR 為 0.27；24 個月年齡層小於等於 4 分者 LR 都是無限大，5 分者(1 題不通過)LR 為 1.73，6 分者 LR 為 0.35。綜合各年齡層之多層次概率比，其不同分數之 LR 值，略有不同；5-6 分者(0-1 題不通過者)，其 LR 皆小於 1；3-5 分者(1-3 題不通過者)其 LR 皆介於 1-10；≤ 2-4 分者(2-4 題以上不通過者)其 LR 皆為無限大，符合 SpPin 之標準。

五. 已知族群效度

如表六所示，三個年齡層一般兒童的 DICHP 總分皆顯著地高於遲緩兒童 (Mann-Whitney $U = 367-1042$; $p < 0.001$)，因此 DICHP 初修版具已知族群效度。

討論

修改部份文字後，12 至 36 個月 3 個年齡層之 DICHP 初修版之內部一致性與再測信度可；以 ROC 分析，篩選正確率為 0.82-0.86；若根據最佳切截點，3 個年齡層之敏感度為 0.69-0.80，特異度為 0.76-0.89。陽性概率比範圍為 3.3-6.3，因此測試結果若為陽性，其診斷影響力輕度有時重要至中度；陰性概率比 0.26-0.35，因此測試結果若為陰性，診斷影響力輕度有時重要。其也具已知族群效

表五：3 個年齡層「兒童健康手冊」發展題目初修版之多層次概率比

總分	一般兒童 (n)	遲緩兒童(n)	概率比	診斷影響力
12 個月				
0	0	2	∞	SpPin ¹ , 高度
1	0	4	∞	SpPin ¹ , 高度
2	0	3	∞	SpPin ¹ , 高度
3	6	5	2.19	輕度有時重要
4	4	11	7.23	中度
5	24	7	0.77	輕度重要性少
6	58	3	0.14	中度
總計	92	35	—	—
18 個月				
0	0	7	∞	SpPin ¹ , 高度
1	0	3	∞	SpPin ¹ , 高度
2	0	4	∞	SpPin ¹ , 高度
3	0	5	∞	SpPin ¹ , 高度
4	2	4	2.27	輕度有時重要
5	10	12	1.36	輕度重要性少
6	38	9	0.27	輕度有時重要
總計	50	44	—	—
24 個月				
0	0	9	∞	SpPin ¹ , 高度
1	0	2	∞	SpPin ¹ , 高度
2	0	4	∞	SpPin ¹ , 高度
3	0	16	∞	SpPin ¹ , 高度
4	0	13	∞	SpPin ¹ , 高度
5	7	16	1.73	輕度重要性少
6	59	27	0.35	輕度有時重要
總計	66	87	—	—

¹ SpPin：確定陽性診斷

表六：「兒童健康手冊」發展題目初修版發展遲緩兒童與一般兒童於各年齡層總分之差別

	遲緩兒童			一般兒童			U(p)值 ¹
	人數	中位數	範圍	人數	中位數	範圍	
12 個月	35	4	0-6	92	6	3-6	446.0(<.001)
18 個月	44	4	0-6	50	6	4-6	367.0(<.001)
24 個月	87	4	0-6	66	6	5-6	1041.5(<.001)

¹ Mann-Whitney 檢定；p 值(probability values)

表七：3 個年齡層「兒童健康手冊」發展題目初修版篩選結果臨床運用之建議

年齡組	結果/建議	正常無介入	二篩	診斷測試
12 個月 A		5-6 題通過	2-3 題不通過	≥ 4 題不通過
12 個月 B		全通過	1-3 題不通過	≥ 4 題不通過
18 個月		全通過	1-2 題不通過	≥ 3 題不通過
24 個月		全通過	1 題不通過	≥ 2 題不通過

度。本研究之優點為收集資料完全仿臨床之流程，由兒童之家長或主要照顧者於健兒門診場域中填寫，因此研究結果可推論至臨床實務之運用。由本研究結果來看，12、18、24 個月 DICHP 分量表具可接受之信效度，應可配合幼兒之預防保健，作為發展遲緩兒童早期發現之監測工具。

由於 DICHP 並未具常模資料，因此只能稱之為發展監測工具，不能稱之為發展篩選工具[10]；然即使是發展監測工具，於臨床使用仍須具備基本效度特質，因此將其與過去發展篩選工具之效度資料比較。「丹佛兒童發展量表」敏感度為 0.68、特異度 0.92[11]；Shoemaker 等人以「貝萊氏嬰兒發展量表」為標準，顯示「明尼蘇達兒童發展量表」的敏感度 0.56 與特異度 0.92[12]；以「嬰幼兒綜合發展測驗診斷量表」為黃金標準，具有台灣常模之「嬰幼兒綜合發展測驗篩選量表」於最佳切截點下，於 5-18 月嬰兒期其敏感度為 0.89，特異度 0.86；於 19-42 月大幼兒之敏感度為 1.0，特異度 0.85；台灣學者編製之「簡易兒童發展量表」於 5-18 月大嬰兒期其敏感度為 0.44，特異度 0.97[13]。以醫學診斷與特教老師綜合研判結果為效標，「嬰幼兒綜合發展測驗篩選量表」於 3 至 71 個月之幼兒所得之敏感度為 0.78，特異度 0.89[14]；因此與過去篩選工具比較，DICHP 初修版之篩選效度尚不差，所以可當第一線之發展監測工具之一。

據研究者所知，本研究是第一篇使用多層次概率比分析兒童發展監測工具之文獻。如前所言，測量工具其提供多層次概率比之參考價值高於單截點。根據本研究 ROC 分析與多層次概率比結果，作者對 3 個年齡層 DICHP 初修版其篩選結果之臨床運用建議如表七所示。於 12 個月大年齡層，有 AB 兩種策略，根據多層次概率比結果，<3 分者 (>=4 題不通過)，符合陽性結果可確定診斷(SpPin)原則，應進一步轉介診斷測試；依 ROC 分析，最佳切截點是 5，因此 ≥5 分者(即 5-6 題通過)可視為正常；3-4 分者(1-2 題不通過)，因其 LR 在 2.19-7.23，皆大於 1，在此分數範圍其測試後發展遲緩機率會高於測試前機率，然診斷影響力屬於輕度至中度，因此建議應對這些兒童進一步二篩，此即 A 策略。此外，12 個月 B 策略在選取正常之切截點標準完全根據多層次概率比之值，因為 5 分者

其 LR 為 0.77，診斷影響力程度為輕度重要性少，因此建議 3-5 分者(1-3 題不通過)，進一步二篩；6 分者(全通過) LR 0.14，屬於中度診斷影響力，可判斷為正常。18 個月與 24 個月年齡組其 ROC 與多層次概率比分析結果皆建議 6 分者(全通過)才視為正常。於 18 個月大，根據多層次概率比結果，<4 分(>=3 題不通過)，符合陽性結果可確定診斷(SpPin)原則，應進一步診斷測試；4-5 分(1-2 題不通過)，進一步二篩。於 24 個月大，根據多層次概率比結果，<5 分(>=2 題不通過)，符合 SpPin 原則，應進一步診斷測試；5 分(1 題不通過)，進一步二篩。

測量工具之效度檢驗，通常可分為兩方面，即測量之效度(validity of measurement)與決策之效度(validity for decision)。本研究偏向決策之效度，即分析 AUC、LR、敏感度與特異度，以供臨床運用之參考。測量之效度則屬傳統之效度，包括內容效度、同時效度、預測效度與建構效度。DICHP 初修版修改之時，已邀各專家進行題目審視，因此具有內容效度；本研究所提供之已知族群效度可視為建構效度之一種；而同時效度與預測效度則須與黃金標準工具測試結果進行相關分析，這兩種效度尚待後續研究。此外，依據診斷測試工具研究之內在效度七個方法學標準(7 methodology standards for diagnosis test study)[15]，本研究符合 6 個標準，即(1)樣本構成內容標準(spectrum composition)，提供樣本年齡和性別分佈，包括收取樣本的條件及臨床特徵；(2)恰當之小組分析標準(analysis of pertinent subgroups)，針對不同年齡層樣本進行個別分析；(3)測試準確性之結果的精準度標準(precision of results for test accuracy)：AUC 值提供 95%信賴區間值；(4)測試重現度標準(test reproducibility)：即所用發展測試具再測信度；(5)診斷偏差之避免標準(avoidance of workup bias)：即若黃金標準為醫師診斷，計算分數者並不知醫師之診斷；(6)複檢偏差之避免標準(avoidance of review bias)：如醫師診斷時並不知此篩選測驗分數。但由於本研究是病例對照研究，無法符合其他 1 個標準，即不確定之測試結果的呈現標準(presentation of indeterminate test results)；因此總結本研究在內在效度為可接受範圍。

發展篩選(developmental screening)之最大問題是：即使高效度之篩選工具，發展遲緩低盛行率常導致低陽性預測率[16]，多階段發展偵測(multiphase surveillance)是解決方法之一[7, 16]；美國小兒科醫學會因此提出發展遲緩早期發現建議流程為定期篩檢(periodical screening)[10]，主要是考慮發展病理的變動性，藉由連續性追蹤篩檢提高篩檢的效度，以減少成本。近年來之研究亦顯示，若能將父母對兒童發展之主觀感覺放入，將有助於篩檢之準確率與降低篩檢成本，並適合忙碌的小兒門診[17-18]。然而國內目前只有個別篩選工具之信效度資料，宜進一步探討國內現有發展篩選如何組合成多階段發展偵測，以具最有效度及最有成本效益。

根據 Muir Gray 之建議[16]，發展篩選工具信效度研究僅是全國性施行全面篩檢測驗之第一步，篩檢測驗之運用須經 8 個階段，方可於全國施用並達良好效用。各階段如下：(1)研究；(2)衛生主管機關研擬篩檢流程；(3)技術評估；(4)成立政策顧問委員會；(5)全國政策訂定，如是否全國施行、基於地區需求同意地區之評估決定、提出不宜使用之測驗；(6)協調全國落實篩檢之進行；(7)全國篩檢網絡之建立；(8)持續改善全國篩檢計畫之品質。因此台灣有必要成立發展篩選政策顧問委員會類似組織。

本研究之限制如下：本研究之效度研究為病例對照研究，由於病例對照組研究之患童通常較屬中重度，會高估診斷正確率[19]，因此參考本研究結果宜採較保守態度；建議未來應再進行更大型之連續性個案研究，以確認其於臨床使用之效度。此外，本研究只針對 3 個年齡層，其他年齡層之信效度研究宜在未來進一步研究。最後，本研究以方便取樣收取個案，大部分兒童來自台灣北部地區，以大專教育程度及台籍父母居多，雖有南部、非台籍及國中教育程度之父母，在研究結果之推論上較適用於都會區與母親教育程度大專以上者，建議在其他場域，尤其是低社經族群與新入籍父母，其信效研究仍需進一步進行。此外，由於各年齡層一般兒童與遲緩兒童之比例不同，如 24 個月年齡層一般兒童與遲緩兒童之比例約為 1:1，6 個月年齡層其比例則為 3:1；因此造成整體遲緩兒童之平均年

齡高於一般兒童；然若分別比較個別年齡層之兩組兒童年齡，並無顯著差別；因此年齡差別因素不會影響本研究信效度結果。

總結來說，本研究之 3 個年齡層 DICHIP 之信效度尚可，可當多階段篩選之第一層工具之一。根據多層次概率比，建議 $\geq 2-4$ 題不通過，應轉介診斷測試；1-3 題不通過進一步二篩；5-6 題通過，於該年齡階段視為正常。然而評估兒童原則是多訊息來源、週期性，發展監測工具僅是多種評估方法之一，於臨床尚可根據觀察、家長提供之資訊與身體檢查等資訊調整測試前機率[20]。

致謝

本研究為行政院衛生署研究計畫，計畫編號：DOH95-HP-1205。感謝所有參與此研究之家長兒童，及協助收集個案單位：台大醫院小兒部、復健部，台北榮總復健部，台北聯合醫院婦幼院區，臺北市早期療育綜合服務中心，台北市新光醫院復健部，嘉義莊凱全小兒科診所，台北市社會局。提供意見之專家學者：朱曉慧醫師、邱南昌主任、周怡宏主任、林幸君主任、林美瑗秘書長、洪焜隆院長、洪百薰簡技、陳質采醫師、遲景上主任、戴政教授(按筆畫順序)；研究人員：郭郁伶、王湘慧、陳佳琳、廖雅慧、林欣儀、劉妍伶。

參考文獻

1. American Academy of Pediatrics, Committee on Children with Disabilities: Developmental surveillance and screening of infants and young children. *Pediatrics* 2001;108:192-6.
2. Cochrane AL, Holland WW: Validation of screening procedures. *Br Med Bull* 1971; 27:3-8.
3. 黃月桂(長庚大學，長庚醫學暨管理學院)。全民健保預防保健服務之利用與成本效益分析。期末報告。行政院衛生署；1997 年 12 月 31 日。計畫編號：DOH86-NH-009。
4. 吳雪玉、廖華芳、姚開屏、李旺祚、王天苗、謝正宜：「嬰幼兒綜合發展測驗」動作分測驗與「皮巴迪動作發展量表第二版」的診斷

- 準確度。台灣醫學 2005;9:312-22。
5. Domholdt E. Physical therapy research: principles and applications. 2nd ed. Philadelphia: W. B. Saunders Company; 2000. p.180-4, p231-4.
 6. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143:29-36.
 7. Straus SE, Richardson WS, Glasziou P, et al: Evidence-based medicine, how to practice and teach EBM. 3rd ed. Edinburgh: Churchill Livingstone; 2005. p.67-99.
 8. Jaeschke R, Guyatt G.H, Sackett DL: Users' guides to the medical literature, III: How to use an article about a diagnostic test, B: What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; 271:703-7.
 9. Finch E, Brooks D, Stratford PW, Mayo NE: Physical rehabilitation outcome measures. 2nd ed. Hamilton, Ontario: BC Decker Inc.; 2002.
 10. Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, and Medical Home Initiatives for Children With Special Needs Project Advisory Committee: Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics* 2006;118: 405-20.
 11. Borowitz KC, Glascoe FP: Sensitivity of the Denver Developmental Screening Test in speech and language screening. *Pediatrics* 1986;78:1075-8.
 12. Shoemaker OS, Saylor CF, Erickson MT: Concurrent validity of the Minnesota Child Developmental Inventory with high-risk infants. *J Pediatr Psychol* 1993;18:377-88.
 13. 廖華芳(台灣大學，物理治療學研究所)。三種發展遲緩兒童早期篩選評估工具之探究。期末報告。行政院國家科學委員會；2006年10月23日。計畫編號：NSC 94-2314-B-002-020。
 14. 王天苗、廖華芳：「嬰幼兒綜合發展測驗」診斷測驗與篩選測驗之判定準確度及切截點。特殊教育研究學刊 2007;30:1-15。
 15. Reid MC, Lachs MS, Feinstein AR: Use of methodological standards in diagnostic test research: getting better but still not good. *JAMA* 1995;274:645-51.
 16. Muir Gray JA: Testing a test. <http://www.jr2.ox.ac.uk/bandolier/band3/b3-1.html>
 17. Squires J, Nickel RE, Eisert D: Early detection of developmental problems: strategies for monitoring young children in the practice setting. *J Dev Behav Pediatr* 1996;17:420-7.
 18. Glascoe FP: Parents' concerns about children's development: prescreening technique or screening test? *Pediatrics* 1997;99:522-8.
 19. Lijmer JG, Mol BW, Heisterkamp S, et al: Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282:1061-6.
 20. Meisels SJ, Atkins-Burnett S: The elements of early childhood assessment. In: Shonkoff JP, Meisels SJ, editor. Handbook of early childhood intervention. Cambridge: Cambridge University Press; 2000. p.231-57.



The Reliability and Validity of the Developmental Items of “Child Health Pamphlet”(DICHP)

Hua-Fang Liao, Ling-Yee Cheng¹, Wu-Shien Hsieh², Ming-Ching Yang³,
Kou-Su Tsou⁴, Kuen-Ying Tsai⁵

Abstract: The developmental surveillance used for the early detection of developmental delay for children has been conducted for many years in Taiwan. However, the information of reliability and validity of the developmental surveillance items in the Child Health Pamphlet (DICHP) edited by the National Health Bureau of the Department of Health were still not very clear. The purpose of this study was to investigate the reliability and validity of the DICHP. The methodology research was used for reliability study, and the case-control study was used for validity study. A convenient sampling method was used to recruit 395 children (developing typically: children with developmental delay =1-3: 1) in 3 age groups from 12- to 36-month-of-age, among them 194 children tested 2 times within 1 week for test-retest reliability. Before data collection, the developmental items were inspected by the research group, and a wording revised DICHP (6 items for each age group) was used in this study. The parents or caregivers filled the 6 items according the child's performance. Then the area under the receiver operating characteristic (ROC) curve, multi-level LRs, and the known-groups validities were analyzed. The results showed that the revised DICHP have acceptable internal consistency ($\alpha=0.74-0.86$) and test-retest reliabilities ($r_s=0.72-1.00$, $p<0.001$). The screening accuracy were with area under ROC of 0.82-0.86 ($p<0.001$). The sensitivities were 0.69 -0.80, specificities were 0.76-0.89. The LRs of the DICHP scores of 5-6 were less than 1, of $\leq 2-4$ were infinity, meeting the SpPin rule. It has acceptable known group validity. The authors concluded that for children aged 1-3 years, the DICHP had acceptable reliabilities and validities, and could be used as the first tool for multi-phasic surveillance. From the ROC and multi-level LRs of the DICHP, authors proposed that further diagnosis test was recommended for children with more than 2-4 failure items, second screening test for children with 1-3 failure items, and no intervention for children with 0-1 failure item.

Key Words: developmental delay, screening test, reliability and validity, likelihood ratio, child

(Full text in Chinese: Formosan J Med 2008;12:502-12)

School and Graduate Institute of Physical Therapy, College of Medicine, National Taiwan University; Department of Rehabilitation Medicine, National Taiwan University Hospital; ¹Department of Rehabilitation Medicine, Taipei Veterans General Hospital; ²Department of Pediatrics, National Taiwan University Hospital and National Taiwan University, College of Medicine; ³Graduate Institute of Health Care Organization Administration, College of Public Health, National Taiwan University ; ⁴Department of Child Psychology, Taipei Municipal Women's and Child's Hospital; ⁵Department of Special Education, Taipei Municipal University of Education, Taipei, Taiwan

Address correspondence to: Hua-Fan Liao, No.17, Xushou Rd., Taipei, Taiwan