

由資料品質談家庭收支調查 在社福議題的運用

呂朝賢

壹、前言

運用社會調查資料分析社會問題已成當今社會科學的常態，若以與社會福利相關的社會調查歷史來說，其實運用社會調查來探索及分析社會問題的歷史並不長，自 Charles Booth（註 1）於 1886 開始調查倫敦的勞工生活狀況，並於 1889 至 1903 年間陸續出版的 17 冊 *Life and Labour of the People of London* 算起，迄今也才一百餘年而已。要是再考究資料調查品質、嚴謹度與調查方法，如：隨機抽樣技術的運用（若以 Sir Arthur Bowley & A.R. Burnett-Hurst 於 1915 年版的 *Livelihood and Poverty: a study in the economic conditions of working-class households in Northampton* 算起）則恐怕真正的歷史是不及一百年的。

在過去的一百年中，運用社會調查分析社會問題之所以日益普及，除了與調查有關的統計知識與技術發展日趨成熟有關外，電腦科技的發達降低資料流通與運用的障礙亦是關鍵因素。但隨著社會問題日趨複雜、社會調查的種類與社會調查的內

容日益多元、社會調查資料總落後於社會問題的發生……等條件之下，如何適切運用既有的社會調查資料探索已知與未知的社會議題，已成為社會科學界關懷的重點之一。

由於臺灣較缺乏「補助」或「支持」學術研究的基金會與私人捐贈，因此與臺灣有關的大型的或持續性的社會調查通常是由政府資金贊助或補助進行的。雖然許多政府所進行的調查內容常需配合行政需要而做，無法全然依理論構念、研究者需求、社會問題類型而設計；但對於民間學術資源不豐富的社會福利學界，這些政府所進行的社會性調查資料還是彌足珍貴。例如：內政部所做的「臺閩地區老人狀況調查」、「臺閩地區少年身心狀況調查」、「臺閩地區國民生活狀況調查」、「臺閩地區身心障礙者生活需求調查」、「臺閩地區低收入戶生活狀況調查」等十餘類與社會福利相關的調查，已成為分析我國社會福利相關議題的重要次級參考資料來源。（註 2）

本文擬引介的資料為行政院主計處

「家庭收支調查」(註 3)，該調查源自臺灣省政府 1954 年的「薪資階級家庭調查」，於 1959 和 1963 年則以「都市消費者家計調查」之名各調查一次，1964 年擴大成全省性的家庭收支調查，1966 年改為每兩年調查一次，1970 年改按年辦理，在 1970 年、1979 年因政府組織更改，主持調查的單位由原來的臺灣省政府，再增加臺北市政府及高雄市政府。1999 年因精省，原省府辦理部分，改移由行政院主計處主辦，目前由北、高二市主計處及行政院主計處按年共同辦理調查事宜，臺灣地區資料之彙總、編製報告則由行政院主計處中部辦公室負責。(行政院主計處，2005a、2005b) 該份社會調查不僅歷史悠久，其調查內容亦已成許多社會問題剖析的重要原始資料來源。該份資料已被收錄於盧森堡所得調查研究 (The Luxembourg Income Survey Study) 中，成為極少數國際學術研究者容易取得的臺灣資料，為國際上比較所得分配與貧窮的重要資料來源。(註 4) 這些理由，皆是本文選擇以該資料做為引介對象之因。

雖然調查研究可能是社會科學家收集原始資料最經濟的作法，(Babbie, 2001:238) 但這不意謂調查研究所得的資料皆具備良好品質。調查資料從抽樣架構 (sampling frames) 建立至電子資料建檔完成的一系列過程中，皆可能會產生誤差，各種誤差 (註 5) 的多寡會直接影響調查資料品質。但何謂資料品質 (Data Quality)？事實上並無一致性的定義，傳統上，統計學家通常認定資料品質就是指

資料的正確性 (accuracy)，調查者透過比較樣本與母體間的偏誤 (bias)、變異數 (variance) 與平均誤 (mean squared error) 等統計量來衡量資料的品質。(Groves, 1987) 但此狹義的界定卻無法完全表達資料品質，這個多面向的概念，尤其是在不同調查時間、人力與經費的限制下，不同的領域的研究者對於資料品質的要求面向並不盡相同。(Stoop et al., 2002)

基於此，本文對資料品質採用較寬的界定，我們擬從相關性 (Relevance)、正確性 (Accuracy)、適時性 (Timeliness)、連貫性 (Coherence)、可近性 (accessibility)、可詮釋性 (Interpretability) 等六個品質組成向度，(Statistics Canada, 2003:85-92；Brackstone, 2000) 來檢視家庭收支調查資料對於社會福利相關議題分析的潛力。

貳、相關性

「相關性」係指調查資料的內容或所可提供的資訊，是否契合研究者所需，足以回答研究問題。因為係依資料對研究問題／研究者的契合度來評價，故同一筆資料對於不同研究者的重要性必然是不同的。(Statistics Canada, 2003:85-92；Brackstone, 2000) 如行政院主計處所述：家庭收支調查項目包括家庭戶口組成、家庭設備及住宅概況及所得收支與消費支出，據以計算家庭設備普及率、自有住宅率及平均每戶可支配所得、消費及儲蓄。(行政院主計處，2005b)

因此，與家庭有關的收入與支出大多

可由家庭收支調查中得知。與其它國家一樣，家庭收支調查亦被運用在分析臺灣的貧窮問題；而這亦是臺灣社福界運用該份資料最普遍的研究類型。除此之外，筆者認為該筆資料還可運用在如下幾種社會福利議題上：

民眾的社會福利受益程度與福利受益的分配議題：隨著我國社會福利政策的轉變，自 1990 年代起各項福利津貼項目不斷被政府提出，家庭收支調查單位似乎亦注意到此議題，於是自 1994 起陸續加入相關的問項，例如：1994 年起加入了「低收入生活補助」、「老人津貼」、「殘障生活補助」、「災害急難救助等其它補助」；1995 年起則加入「老農年金」、「補助全民健保保費」。這些資料若再搭配相關社會保險支出項目及家戶的基本資料，將可有助於我們分析不同人口特徵、不同社會階層及不同家庭型態的福利受益狀況。

各類福利可能受益者的需求成本估算：由於家庭收支調查對家庭的支出項目有相當細緻的調查，社會福利研究者可依此計算諸如：老人基本生活費用、兒童養育成本、低收入者的最低生活費等具政策意涵的數據（事實上，我國社會救助法中，最低生活費的界定，即是依據該筆資料結果）。以作為未來政府相關福利給付調整之參考。相同的，我們亦可依現行各類人口所得之福利給付項目的多寡，來評估現行福利措施的效率與效益問題，或更進一步討論整體與區域間社會福利資源分配性公正的議題。

除了以上二種直接與社會福利措施有

關的議題之外，我們亦可以運用家庭收支調查分析諸如：家庭勞動人口組成與型態、家庭生活型態、家庭型態的轉變、家庭支持系統、家庭資源的分配等等較屬基礎性的社會福利研究議題。或者，若能得到政府相關單位的適當配合的話，我們亦可以家庭收支調查的樣本當做其它社會福利性調查的樣本，如此，不僅可以使該筆資料的運用更深廣泛，亦可以使其它與社會福利有關的調查成本下降，但內容卻更豐富。

參、正確性

「正確性」係指調查所得的統計訊息與所要描述的現象相符的程度，或說真實地反映所欲測量現象的程度。傳統上，係以統計方法來估量正確性，造成準確性的來源可區分為二：1. 偏誤（bias）／系統性誤差（systematic error），2. 變異（variance）／隨機誤差（random error）。（Statistics Canada, 2003:85-92; Brackstone, 2000）

家庭收支調查係以「戶籍登記資料以縣市為副母體，採分層二段隨機抽樣方法，抽出受訪戶，以就業人口產業結構比例將各村里分為都市、城鎮及鄉村三層（臺北市及高雄市除外）；第一段抽樣單位為村里，第二段抽樣單位為戶。每年由調查員訪問一次，查詢其全年所得收支主要項目，以戶數權數加權得到臺灣地區資料，其中戶數權數係以訪問戶代表該村里戶數乘以該村里代表所在層母體戶數。」（行政院主計處，2005b）而為確保資料的正確

性，主計處將「收支科目及消費科目做極端值及交叉檢查，結果表並按區域別、都市化程度別、職業別等交叉分類，確保資料正確性。」（行政院主計處，2005b）再者，該份調查包含二類受訪戶，其中：「訪問調查戶，每年由調查員訪問一次，查詢其全年所得收支主要項目。記帳調查戶，除接受訪問調查外，每日並根據家庭實際收支逐筆記帳，調查員須按日檢查帳簿，以防記載錯誤與遺漏。訪問調查，因受記憶誤差與遺漏之限制，較記帳調查之正確性為小，但因實施記帳調查所需人力、經費均較訪問調查為鉅，難以大量實施，故對部分家庭實施二種調查，比較兩種調查結果之關係，以校正全體訪問調查之結果。」（行政院主計處，2004）另外，在每年主計處皆會出版調查的報告書，書中除相關統計結果外，對於調查有關的統計方法、程序與誤差的估計皆有清楚的交待。

家庭收支調查雖有如上的嚴謹調查程序，但仍難免會有誤差產生。主要的問題來源如下：1.膨脹係數設算不適當，以致歷年來在總人口的推估上皆有偏誤，高出公務統計甚多；2.拒訪率逐年提高，所採用的備用戶屬性又偏離原樣本，衝擊調查的代表性；3.記帳戶不易找尋，且人數過少，紀錄困難。4.調查員找尋不易、訓練困難與流動過大，可能影響調查品質；5.調查內容過於複雜，記帳內容設算太多，不易估算且易成偏誤；6.記帳內容與調查內容時間相差距一年，無法有效配合。（蔡鈺泰，2001；詹德松，1987；賴美齡，1986）

上面所羅列的各項問題有些仍存在，

但有些則已在近年來得到適當的解決。這些問題皆對資料的正確性產生一定程度的影響，例如：家庭收支調查所呈顯的低收入人數比例就與實際的低收入人數比例有些微落差；某些社會邊緣群體如：遊民、受機構安置民眾、無戶籍人口等在這該調查中是無法被抽出的；又如某些年度家戶的可支配所得可能是負數的狀況；或者戶量過大，數值不太合理。這些問題在愈早的調查年度中愈易被發現。另外，家庭收支調查所界定的「家庭」定義與一般社會福利法令中的「家庭」定義，並不完全相同，亦無法從資料的操作上調合彼此的差異，這也是在運用分析時應該注意的資料限制。

雖然家庭收支調查有如上的瑕疵，但尚在可容忍的範圍中，多數研究者亦對該筆資料正確性持肯定的態度。這些資料數據的錯誤，是提醒欲運用該筆資料的社會福利界研究者，尤其是欲分析某項議題的長期趨勢時，應當對資料做清理與檢查，甚至是重新填補（imputation）與替代不適當原始數據。如此，方能更正確地描繪所欲了解的社會事實。

肆、適時性

「適時性」指資料收集時的參考時間點與所欲知訊息間的時間差距，或與電子資料可被使用者使用間的時間差距。資料的適時性會影響資料的相關性，（Statistics Canada, 2003:85-92；Brackstone, 2000）若電子資料可使用的時間與收集資料所界定的參考時間差距過大，或者資料收集的參

考時間點與問卷中所欲回答的問題時間差異過大（例如：問成人，回憶其兒時的父母教養態度），皆會降低資料的適時性與相關性。

通常家庭收支調查從開始調查至資料釋出約需 1 年左右時間，若再考慮資料採用時間點則皆在 2 年以內。就資料釋出的時程來說，該筆資料的適時性相當高。比較有問題的部分是，該調查的調查戶所詢問的內容皆是調查前一年家戶的收支狀況，又詢問的收支項目相當細緻，甚至有些收支項目若受訪者沒有記帳的習慣，則可能依自己「設算」的結果填答，調查結果因而會受受訪者記憶所影響，而影響其正確性與適時性。這一部分的問題，雖然可以經由偏離值、記帳戶資料的調整來修正，但依然未能完全解決。所以，有些研究者建議調整調查的時期或頻率，來降低此問題所造成的資料偏誤。（詹德松，1987；賴美齡，1986）

不過，對於社會福利相關研究議題而言，筆者認為家庭收支調查資料的適時性瑕疵，並不是很嚴重，尤其我國的社會福利的受益者的評定通常是以年為基礎，與家庭收支調查的調查時程相仿，應不致受到受訪者的記憶之影響而造成很大的偏誤。而即便給付不以年為基礎，我國相關福利給付額度皆相當固定、給付項目亦不多，因此受訪者應可比較正確地提供實際的社會保險等福利收入與支出額度，而不致受到記憶太大的影響。另外，因福利給付額度與支出額度皆較固定，若遇到受訪者所填答的數值太過極端時，亦可以做很

好的補正。

伍、連貫性

「連貫性」係指在一個廣泛的分析架構或時間上，調查所取得的資料與其它統計資料相結合、互補的程度。當然若是用標準化的概念、分類與目標人口群，則調查資料的連貫性就會較佳。（Statistics Canada, 2003:85-92；Brackstone, 2000）

由於家庭收支調查原初的調查目的並非針對社會福利相關研究議題而做，因此，並無法與其它內政部調查資料直接做連結，除非使用的樣本是同樣的才有可能。但就家庭收支調查本身來說，其歷年 37 次的調查，調查的一致性是不錯的，因此可以做為諸如：貧窮趨勢、家庭型態趨勢等議題的分析資料來源。另外，該調查所得的收入與支出結果，也可以做為比對或補充其它社會福利相關調查的不足，如：老人生活狀況調查、低收入生活狀況調查、身心障礙者需求調查等。

陸、可近性

「可近性」是指調查資料訊息，是否提供了適當的調查過程資訊、初步統計數據、儲媒、傳遞方式及資料讀取程式等，幫助使用者知曉調查資料與自己的研究議題間的配適程度、取用成本與限制為何，並進而能設計如何降低資料取用限制的方法。（Statistics Canada, 2003:85-92；Brackstone, 2000）

家庭收支調查是一份歷史悠久的政府調查統計，目前 1976 年至 2003 年的原始

資料檔已為中央研究院「學術調查研究資料庫」所收錄，該資料庫提供學術研究者線上資料申請功能，所提供的資料包括：過錄編碼簿、問卷、調查基本資料表、讀取資料程式及原始電子資料；是一份可近性相當高的資料。

大多數的研究者係以 SAS 統計套裝軟體來讀取家庭收支調查電子資料檔，主計處與學術調查研究資料庫所提供的範例程式亦是以 SAS 語法撰寫的（但前者所提供的範例程式對初學者太精簡了，後者的程式則又過於繁瑣）。筆者猜想這可能與該筆資料部分內容係以 EBCDIC（Extended Binary-Coded Decimal Interchange Code）編碼所致，該種編碼格式是 IBM 所發展出來的，曾廣泛運用在 IBM 電腦上，但對於終端的使用者，其實多數並不清楚／也不會想去瞭解主計處為何要採用這類編碼型式（註 6），而非一般 PC 常用的 ASCII（American Standard for Information Interchange，美國資訊交換標準碼）。當然，若研究者不了解這種字元編碼系統／型態，只要會寫 SAS 程式語法或任何看得懂 EBCDIC 碼的軟體即可。問題就出在這裡，多數社會福利相關科系的統計教學軟體，多以 SPSS 或 EXCEL 為主要教學工具，以致許多研究生或研究者在實際運用該筆資料時，需花費一段很長的時間學習 SAS，以確保不抓錯資料。當然，以上所提之問題完完全全不是調查單位或資料提供單位的錯，是研究者本身能力有限所致，但如果由全面品質（Total Quality Management）的向度來看，則或許資料提

供單位可以考慮加入較友善／簡單的讀取該筆資料的方式，以縮短使用者與資料間的距離。

柒、可詮釋性

「可詮釋性」指調查資料是否提供充足的訊息諸如：資料編碼簿、調查範圍、調查時間、調查對象、名詞定義與解釋等，幫助使用者正確的詮釋資料內容，做適當的分析。（Statistics Canada, 2003:85-92；Brackstone, 2000）

自 1976 年算起，家庭收支調查已進行過 37 次的調查，雖然在問卷內容上曾有多次更動，但歷年來主計處所提供的調查背景、對象與名詞定義與解釋皆相當一致且清楚。但在資料編碼簿上就比較缺乏，在早期主計處所提供的供外界分析的資料編碼簿比較不全，但近年來主計處所提供給外界的編碼簿已相當平易近人，並且將歷年各項問項的演變製表羅列，相當有益於使用者了解。目前中央研究院「學術調查研究資料庫」已收錄家庭收支調查，並對問卷、編碼簿重新做更清晰簡易的整理，對於一般未接觸過此資料的使用者而言是相當好的了解該筆資料的指南。

當然這不意謂家庭收支調查沒有改善的空間，在國外的政府調查資料通常經過幾年後就會開始檢討調查的內容，並且提供調查整體調查錯誤（Total Survey Error）的處理方式、設算值的處理方式，並對調查實務程序、內容做檢討，尤其是為何更動調查項目的討論與解釋。這類訊息雖較屬於調查統計領域研究者關懷的議題，主

計處亦有舉辦類似的研討會，但卻甚少直接提供給使用者。筆者以為，若能將此類資訊，提供給使用者，將可降低錯用資料的可能性。

捌、討論與結語

運用既有調查資料分析社會福利議題，已成為學界的常態。本文藉由資料品質的不同面向，討論運用行政院主計家庭收支調查分析社會福利議題的可能性。分析結果顯示，該筆資料在資料品質的 6 個向度上，皆具有中高以上的水準，相當適合做為社會福利研究者分析相關議題的參考資料。當然這不意謂，該筆資料是完美無缺的，事實上，每一筆既有的調查資料皆有其適用範圍的限制；或者說，從既有的統計資料中，我們大多只能分析所欲分析的「社會事實」的部分而已，但學術是一種累積的過程，所有研究議題不可能在資料完美的狀況下再去探索；若是如此，這無疑如同痴人望天，整天張目瞪視老天，祈求可以掉下禮物般，一樣無知。既

有的次級資料或許不完美，但知識是累積的，社會事實的描繪如同考古一樣，需一點一滴拼湊、聚合起來。當您在古蹟中挖到一片瓦時，不要因為它不是一棟屋子，就急著將它丟掉，您應該先分析、記錄它，然後也許您就可因此而發現一個被我們所忽略的新世界。

家庭收支調查原初的調查目的雖非為研究社會福利相關議題而設的，但基於該筆資料的資料品質向度皆處於中高水準以上，筆者認為社會福利研究者，若所欲分析的問題涉及「家庭收入與支出」面向的福利議題時，皆可將該筆資料設為第一優先選用的次級資料，或將其調查統計結果視為探析時的重要參考或用以比對的基礎資料；如此，不僅有助於提高議題分析結果的可信度與效度，對整體社會福利政策所提的政策建議也會因此而更具合理性與實用性。

（本文作者為南華大學非營利事業管理研究所副教授）

註釋

註 1：一般將社會調查起源的桂冠封給 Charles Booth，但事實上在 Booth 之前就已有「社會調查」的研究，有關 Booth 之前的「社會調查」發展請詳見 Bulmer, Martin, Kevin Bales, and Kathryn Kish Sklar (eds.) (1991) *The Social Survey in Historical Perspective, 1880-1940*. Cambridge: Cambridge University Press. 這本書提供非常詳盡的整理與描述，筆者自認無法說明的比此書中的文章更清楚，所以請有興趣的讀者自己參閱。另外，有關社會調查對貧窮知識的發展與演變的影響，亦可請參閱 O'Connor, Alice (2001) *Poverty Knowledge: Social Science, Social Policy, and The Poor in Twentieth U.S. History*. Princeton and Oxford: Princeton University Press.

- 註 2：有關內政部的調查資料相關資訊與初步分析結果，請參閱內政統計資訊網，網址爲：<http://www.moi.gov.tw/stat/>。至於這些調查資料的取得，除向內政部直接申請外，部分調查資料電子檔可向「中央研究院調查研究專題中心」免費申請，詳細規定請參見其網頁，網址爲：<http://srda.sinica.edu.tw/>。
- 註 3：該筆資料同樣可向「中央研究院調查研究專題中心」免費申請。另外，行政院主計處網站亦提供該筆調查資料的調查說明及若干統計分析結果，詳見下列網址：<http://www.dgbas.gov.tw/lp.asp?CtNode=3239&CtUnit=352&BaseDSD=7>。
- 註 4：有關盧森堡所得調查研究資訊請參閱左列網址，<http://www.lisproject.org/>。
- 註 5：請參閱（Statistics Canada, 2003: 58-59）的說明。
- 註 6：也許可能與以前政府單位所使用的電腦皆是 IBM 大型主機系統有關吧！

參考文獻

- 行政院主計處（2004）92 年家庭收支調查報告，臺北：行政院主計處。
- 行政院主計處（2005a）臺灣地區家庭收支調查，上網日期：2005.05.15，取自網際網路，網址爲：<http://win.dgbas.gov.tw/dgbas04/bc3/a281.htm>。
- 行政院主計處（2005b）臺灣地區家庭收支調查，上網日期：2005.05.15，取自網際網路，網址爲：<http://www129.tpg.gov.tw/mbas/doc4/data/comment.doc>。
- 蔡鈺泰（2001）家庭收支調查制度、技術、分析方法及所得分配衡量方式，行政院主計處出國報告書。
- 詹德松（1987）對家庭收支調查改進意見之商榷，中國統計學報 25（6）：11-24。
- 賴美齡（1986）現行家庭收支調查制度之研究與改進，中國統計學報 24（10）：19-49。
- Babbie, Earl（2001）The Practice of Social Research.9th ed.. Wadsworth/Thomson Learning.
- Brackstone, Gordon（2000）Managing Data Quality at Statistics Canada. Paper to be presented at the statistical Quality Seminar, December 2000, Cheju, Korea.
- Groves, Robert M.（1987）Research on Survey Data Quality. The Public Opinion Quarterly, 51: S156-S172.
- Statistics Canada（2003）Statistics Canada Quality Guidelines.4th ed.. Statistics Canada Catalogue no. 12-539-XIE.
- Stoop, Ineke, Roger Jowell, and Peter Mohler（2002）The European Social Survey: One Survey in Two Dozen Countries. Paper to be presented at the International Conference on Improving Surveys Copenhagen, 25-28 August.