What predictors matter: Risk factors for late adolescent outcomes

Elizabeth Wall-Wieler, MSc, Leslie L. Roos, PhD, Dan G. Chateau, PhD, Laura C. Rosella, PhD

ABSTRACT

OBJECTIVES: A life course approach and linked Manitoba data from birth to age 18 were used to facilitate comparing two important outcomes: high school graduation and Attention-Deficit/Hyperactivity Disorder (ADHD). With a common set of variables, we sought to answer the following questions: Do the measures predicting high school graduation differ from those that predict ADHD? Which factors are most important? How well do the models fit each outcome?

METHODS: Administrative data from the Population Health Research Data Repository at the Manitoba Centre for Health Policy were used to conduct one of the strongest observational designs: multilevel modelling of large population (n = 62,739) and sibling (n = 29,444) samples. Variables included are neighbourhood characteristics, measures of family stability, and mental and physical health conditions in childhood and adolescence.

RESULTS: The adverse childhood experiences important for each outcome differ. While family instability and economic adversity more strongly affect failing to graduate from high school, adverse health events in childhood and early adolescence have a greater effect on late adolescent ADHD. The variables included in the model provided excellent accuracy and discrimination.

CONCLUSION: These results offer insights on the role of several family and social variables and can serve as the basis for reliable, valid prediction tools that can identify high-risk individuals. Applying such a tool at the population level would provide insight into the future burden of these outcomes in an entire region or nation and further quantify the burden of risk in the population.

KEY WORDS: Attention deficit disorder with hyperactivity; education; risk; longitudinal studies

 Lebenslaufansatz und gekoppelte Daten aus Manitoba von der Geburt bis zum Alter von 18 wurden verwendet, um folgende Fragen zu beantworten: Die Maßnahmen, die die High School Abschlussquote vorhergesagen, unterscheiden sich von denen, die die ADHD vorhergesagen? Welche Faktoren sind am wichtigsten? Wie gut passen die Modelle zu jedem Ergebnis?

METHODEN: Verwaltungsdaten aus dem Population Health Research Data Repository am Manitoba Centre for Health Policy wurden verwendet, um eines der stärksten Beobachtungsdesigns: multilevel Modellierung von großen Populationen (n = 62.739) und Bruder (n = 29.444) Stichproben. Variableinckommen ergeben sich aus der Verarbeitung der Verwaltungsdaten, die die Familie und Sozialbedingungen im Kindesalter und Jugendalter zwischen 0 und 13 Jahren behandeln. Jede Variable kann als gleichzeitiig stellten für die anderen Maßnahmen. 2-4 Verwaltungsinformationen aus den frühen Lebenslaufansätzen bieten einen besseren Masswert für die üblichen Verhaltensabhängigkeiten von retrospektiven Erhebungen.


Der Einfluss von Adverse Childhood Experiences – einschließlich mangelnder Gesundheit, familiale Instabilität und ökonomische Adversität – auf die Bildung und Gesundheit ist gut dokumentiert. 5-7 Surveys von vielen der Maßnahmen in dieser Arbeit wurden bisher verwendet. 6 Fast alle der Literatur beruhen auf der Ausbildung einer kleineren N und weniger Variablen; die Anzahl von simultanen Kontrollen ist in früheren Studien reduziert worden. Die Forschung auf hoher Schule Abschluss und ADHD typisch bezieht solche gemeinsamen Themen wie Familie Instabilität

Author Affiliations
1. Manitoba Centre for Health Policy, Department of Community Health Sciences, Faculty of Health Sciences, College of Medicine, University of Manitoba, Winnipeg, MB
2. Dalla Lana School of Public Health, University of Toronto, Toronto, ON
Correspondence: Elizabeth Wall-Wieler, Manitoba Centre for Health Policy, Department of Community Health Sciences, Faculty of Health Sciences, College of Medicine, University of Manitoba, 408-727 McDermot Avenue, Winnipeg, MB R3E 3P5, Tel: 204-330-1776, E-mail: wallwiee@myumanitoba.ca
Acknowledgements: This research has been supported by the Social Sciences and Humanities Research Council, the Canadian Institutes of Health Research, and the Canadian Institute for Advanced Research. Data used in this study are from the Health Research Data Repository housed at the Manitoba Centre for Health Policy, University of Manitoba and were derived from data provided by Manitoba Health and Manitoba Education under project #2007/2008-22 and project #2007/2008-20C.
Disclaimer: The results and conclusions are those of the authors and no official endorsement by the Manitoba Centre for Health Policy, Manitoba Health, or other data providers is intended or should be inferred.
Conflict of Interest: None to declare.
and socio-economic disadvantage without allowing comparison across domains. Given the variety of work on high school graduation and ADHD, it makes sense to compare the effects of a uniform set of predictors on these two outcomes. With a common set of variables, we sought to answer the following questions: Do the measures predicting high school graduation differ from those that predict ADHD? Which factors are most important? How well do the models fit each outcome? We use multilevel modelling of large population and sibling samples to answer these questions.

**METHODS**

**Data overview**

The Population Health Research Data Repository at the Manitoba Centre for Health Policy provided routinely collected individual information from all provincial residents. A research registry identifying every individual resident in the province, with arrival and departure dates, deaths and births, was created from the provincial health registry and coordinated with Vital Statistics files. Linked data from several ministries provided information on: individual-level health conditions (Manitoba Health, Healthy Living and Seniors), educational achievement (Manitoba Education), economic adversity (Manitoba Jobs and the Economy), and challenged parents (Family Services). Census data on neighbourhood-level household income, education, unemployment and household family structure were also incorporated.

Children were linked to mothers using hospital birth record information, available in essentially all cases. Both the mother’s encrypted Personal Health Identification Number and the family registration number were used to define siblings.

**Cohort selection**

Our population analysis includes those born in Manitoba, Canada between 1984 and 1989 (inclusive), staying in the province up to June of their 18th year, and included in the Repository. Loss to follow-up (from birth to age 18) was roughly 23% from the original cohort, covering over 98% of those born in Manitoba; out-of-province migration is primarily responsible for this attrition. The original cohort totalled 72,634. Children were excluded for intellectual disability (n = 1,164); missing data on birth weight (n = 1,373); suppressed postal code for children placed with Manitoba Child and Family Services (or Public Trustee) (n = 218); missing family structure history (n = 540); missing Grade 9 Achievement Index score (n = 16); and missing score on the Socio-economic Factor Index (n = 270). Children residing in Manitoba’s Rural North (n = 6,920) were also excluded; many of the schools run by First Nations (Indian) bands in the Rural North do not report graduation figures to the Ministry of Education. Home-schooled children (who do not officially graduate from high school) and those at the few band schools outside the Rural North (n = 2,449) were also not included. Supplementary Appendix A (see ARTICLE TOOLS section on journal site) has additional information on exclusions of individuals from the Rural North and band schools. After exclusions, this sample included 62,739 cohort members.

Sibling analyses (N = 29,444) compare genetically similar individuals growing up in similar environments, correcting for error from omitted parental variables. With siblings born within six years of each other, the family environment will have had a relatively short time to change (in possibly unmeasured ways). Selecting twins in families with twins and other siblings ensured sufficient sample size.

**Variables**

The effect of a series of family- and individual-level covariates on failure to graduate from high school (within four years of entering grade 9) and having an ADHD diagnosis in late adolescence (ages 14–18) is examined (this is not necessarily the first diagnosis of ADHD, but a diagnosis occurred in this time period). Variables are displayed in Table 1. The neighbourhood of the oldest child in the cohort at age 17 is used to determine the Socio-economic Factor Index (SEFI) score and rural residency; all other siblings are likely to live in this neighbourhood as well. Neighbourhood SEFI is closely related to family SES as neighbourhoods are based on six-digit postal codes, which rarely include more than 600 individuals. Some variables examining family situations (such as family structure changes, residential mobility and social assistance usage) are classified as individual-level variables to facilitate sibling analysis. We are examining the presence of these predictors before age 14; one of these events could occur when one sibling is younger than 14 and one older. Several variables are defined by The Johns Hopkins University Adjusted Clinical Group software, which groups 14,000 ICD codes into 32 aggregated diagnostic groups to simplify the large number of possible health measures.

While the presence of family structure changes, residential mobility and health conditions between birth and age 13 is noted, involvement with Child and Family Services (CFS) and income assistance can only be examined in children ages 9 to 13. Data on these covariates were added to the research repository much later than those of the other covariates. Validity issues surrounding mental illness diagnoses using administrative data are discussed elsewhere.

**Procedure**

Observational studies are often criticized for an exaggerated emphasis on particular variables. Stronger research designs and statistical innovations can help to better identify risk factors having causal effects. Our population and sibling analyses use multilevel modelling to both a) handle the lack of independence among observations which violates a key assumption of most parametric procedures, and b) clarify “confounding variables at both within- and between-group levels”. In the multilevel model framework, neighbourhood and family information were placed on the same (second) level, and individual information on the first level. Neighbourhood information was placed at the second (family) level because neighbourhood is not examined at a specific point in time. The neighbourhood of the oldest child at age 17 depends on his/her birth year, and neighbourhoods change. Uniform access to Canadian health insurance further corrected for the possibility of differential insurance coverage within families; the use of objective measures instead of self-reported or retrospective measures strengthens generalizability. Finally, although siblings share many experiences, these experiences may occur at different times in the developmental process. This variation allows using
measures such as residential mobility and family structure changes at the individual level.

Because the same set of predictors and cohort are used to examine two outcomes, adjustments were made for multiple comparisons using the Bonferroni method. A desired confidence level of 95% requires predictors to be significant at \( p < 0.025 \). SAS version 9.3 was used for all analyses.

**RESULTS**

In the population cohort, approximately 20% did not graduate from high school in four years, while 3.2% had an ADHD diagnosis between ages 14 and 18. Table 1 presents frequencies of binary variables and means and standard deviations of continuous measures.

---

**Table 1. Characteristics of population sample**

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Frequencies/percent of population*</th>
<th>Sibling with different outcomes/predictors†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure to graduate high school</td>
<td>12,540 (19.99)</td>
<td>2746</td>
</tr>
<tr>
<td>ADHD, 14–18</td>
<td>1999 (3.19)</td>
<td>736</td>
</tr>
</tbody>
</table>

**Predictors**

<table>
<thead>
<tr>
<th>Family level</th>
<th>Mean standardized SEF2 score (SD)</th>
<th>Rural (25,172 (40.12))</th>
<th>Mean mother’s age at first birth (SD)</th>
<th>Individual level</th>
<th>Male (31,812 (50.71))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean standardized SEF2 score (SD)</td>
<td></td>
<td></td>
<td></td>
<td>Individual level</td>
<td>Male (31,812 (50.71))</td>
</tr>
<tr>
<td>Rural</td>
<td>25,172 (40.12)</td>
<td></td>
<td></td>
<td>Individual level</td>
<td>Male (31,812 (50.71))</td>
</tr>
<tr>
<td>Mean mother’s age at first birth</td>
<td>24.19</td>
<td></td>
<td></td>
<td>Individual level</td>
<td>Male (31,812 (50.71))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Individual level</td>
<td>Male (31,812 (50.71))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Individual level</td>
<td>Male (31,812 (50.71))</td>
</tr>
</tbody>
</table>

**Birth weight, grams**

- Birth weight \( \leq 2500 \) g: 2809 (4.48)
- Birth weight \( 2500 < \) birth weight \( \leq 3500 \) g: 30,914 (49.27)
- Birth weight \( 3500 < \) birth weight: 29,016 (46.25)

**Family size/birth order**

- Only child: 3964 (6.32)
- Two children, first born: 12,272 (19.56)
- Two children, second born: 12,273 (19.56)
- Three children, first born: 6895 (10.99)
- Three children, second born: 7126 (11.36)
- Four children or more, first born: 2705 (4.31)
- Four children or more, second born: 3084 (4.92)
- Four children or more, third born: 3258 (5.19)
- Four children or more, fourth or higher born: 4877 (7.77)

**Family structure history**

- M – parents remain married, ages 0–13: 41,960 (66.88)
- M – parental divorce, no further changes, ages 0–13: 5973 (9.52)
- M – parental death, no further changes, ages 0–13: 649 (1.03)
- M – two or more transitions, ages 0–13: 1502 (2.39)
- U – mother remains unmarried, ages 0–13: 7588 (12.09)
- U – mother marries after birth, ages 0–13: 3724 (5.94)
- U – two or more transitions, ages 0–13: 1343 (2.14)

**Moved between ages 0 and 13: 39,956 (63.69)**

**Health conditions**

- Had ADHD, ages 0–13: 3968 (6.32)
- Had conduct disorders and/or ODD, 0–13: 4365 (6.96)
- Had bipolar disorder, 0–13: 292 (0.47)
- Had asthma, 0–13: 20,078 (32) 5025
- Had major injuries, 0–13: 46,393 (73.95)
- Had other major health conditions, 0–13: 22,390 (35.69)

**Family problems**

- Was a child in care, 9–13: 959 (1.53)
- Received services from CFS, 9–13: 4524 (7.21)
- Economic adversity:
  - Received income assistance, 9–13: 7087 (11.3)
  - Less than average grade 9 achievement: 34,583 (55.12)

---

\* Total number of children = 62,739.

† Total number of individuals with siblings in cohort = 31,504.

‡ Reference category for odds ratios.

**WHAT PREDICTORS MATTER**

Evaluating covariates

Important covariates at birth and during childhood/early adolescence were compared across outcomes. Higher neighbourhood SES decreased the odds of failing to graduate from high school and increased the odds of having an ADHD diagnosis. Rural residency was a protective factor for both outcomes. With increases in mother’s age at first birth, the odds of failing to graduate from high school dropped and those of being diagnosed with ADHD grew. Almost all individual measures (26 of 29) were statistically significant predictors of educational achievement; only 11 individual measures were significant for ADHD. Measures of family size/birth order, family structure history, challenged parenting, and economic adversity were particularly important in predicting high school graduation. Mental health conditions in childhood were the strongest predictors.
predictors of late adolescent ADHD. Tables 2 and 4 present in bold text the odds ratios of the five most important predictors of each outcome.

Presence of the same condition at an earlier age impacted age 14–18 outcomes. For the education outcome, the “same but previous” predictor was “Less than average Grade 9 achievement”; for the ADHD outcome, it was “Had ADHD, Ages 0–13”. Outcomes were modelled without the “same but previous” variable in the models (Model 2 in Tables 2 and 4); in the population models, exclusion of “same but previous” variables increased the importance of family structure changes for both outcomes.

**Model fit**

We used two different statistics to assess model performance. First, a Brier score determined model accuracy; a score of zero indicates complete accuracy and a score of 0.5 indicates complete inaccuracy.20 Second, a c-statistic (ranging from 0.5 to 1) measured model ability to discriminate between those with and those without the outcome. Both models performed very well when all predictors were included (Brier score <0.15 and c-statistic well above 0.8); model fit decreased with removal of the “same but previous” variable, but fit remained reasonable (c-statistic over 0.7) (Table 3).21

### Populations and siblings

Comparisons between population (Table 2) and sibling (Table 4) samples help establish the generalizability of our results through providing somewhat differing perspectives on the life events used. The sibling analysis added the “difference in days between births of siblings” and removed “only child” in the “family size/birth order” measures. With the sample size for the population analyses more than twice as large as that for the sibling analyses, fewer measures in the sibling models were significant (but odds ratios were generally similar). Non-twin siblings were generally less likely to graduate from high school than twins, suggesting higher levels of parental involvement and/or less negative sibling interaction among the relatively small number (n = 1,122) of twins. Siblings born close together (<842 days) and those born far apart (>1,113 days) had the lowest odds of high school graduation.22

### Table 2. Odds ratios of neighbourhood, family and individual covariates (population models)

<table>
<thead>
<tr>
<th>Model</th>
<th>Failure to graduate high school</th>
<th>ADHD, 14–18</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Family level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardized SEFI score</td>
<td>1.46**</td>
<td>1.55**</td>
</tr>
<tr>
<td>Rural</td>
<td>0.71**</td>
<td>0.7**</td>
</tr>
<tr>
<td>Mother's age at first birth</td>
<td>0.95**</td>
<td>0.93**</td>
</tr>
<tr>
<td>Individual level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.4**</td>
<td>1.79**</td>
</tr>
<tr>
<td>Birth weight, grams</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth weight &lt; 2500 g</td>
<td>1.07</td>
<td>1.17*</td>
</tr>
<tr>
<td>2500 g &lt; birth weight ≤ 3500 g</td>
<td>1.11**</td>
<td>1.14**</td>
</tr>
<tr>
<td>Family size/birth order</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only child</td>
<td>1.19*</td>
<td>1.32**</td>
</tr>
<tr>
<td>Two children, second born</td>
<td>1.19**</td>
<td>1.31**</td>
</tr>
<tr>
<td>Three children, first born</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>Three children, second born</td>
<td>1.28**</td>
<td>1.33**</td>
</tr>
<tr>
<td>Three children, third born</td>
<td>1.2**</td>
<td>1.33**</td>
</tr>
<tr>
<td>Four children or more, first born</td>
<td>1.7**</td>
<td>1.72**</td>
</tr>
<tr>
<td>Four children or more, second born</td>
<td>1.73**</td>
<td>1.81**</td>
</tr>
<tr>
<td>Four children or more, third born</td>
<td>1.76**</td>
<td>1.92**</td>
</tr>
<tr>
<td>Four children or more, fourth or higher born</td>
<td>2.16**</td>
<td>2.46**</td>
</tr>
<tr>
<td>Family structure history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M – parental divorce, no further changes, 0–13</td>
<td>1.42**</td>
<td>1.59**</td>
</tr>
<tr>
<td>M – parental death, no further changes, 0–13</td>
<td>1.37*</td>
<td>1.52**</td>
</tr>
<tr>
<td>M – two or more transitions, 0–13</td>
<td>1.52**</td>
<td>1.71**</td>
</tr>
<tr>
<td>U – mother remains unmarried</td>
<td>1.87**</td>
<td>2.18**</td>
</tr>
<tr>
<td>U – mother marries after birth, 0–13</td>
<td>1.57**</td>
<td>1.84**</td>
</tr>
<tr>
<td>U – two or more transitions, 0–13</td>
<td>1.82**</td>
<td>2.07**</td>
</tr>
<tr>
<td>Moved between ages 0 and 13</td>
<td>1.38**</td>
<td>1.43**</td>
</tr>
<tr>
<td>Health conditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Had ADHD, 0–13</td>
<td>1.56**</td>
<td>2.03**</td>
</tr>
<tr>
<td>Had conduct disorder/ODD, 0–13</td>
<td>1.26**</td>
<td>1.36**</td>
</tr>
<tr>
<td>Had bipolar disorder, 0–13</td>
<td>1.78**</td>
<td>2.05**</td>
</tr>
<tr>
<td>Had asthma, 0–13</td>
<td>0.94*</td>
<td>0.94*</td>
</tr>
<tr>
<td>Had major injuries, 0–13</td>
<td>1.08*</td>
<td>1.15**</td>
</tr>
<tr>
<td>Had other major health conditions, 0–13</td>
<td>0.98</td>
<td>1</td>
</tr>
<tr>
<td>Family problems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was a child in care, 9–13</td>
<td>2.11**</td>
<td>2.44**</td>
</tr>
<tr>
<td>Received services from CFS, 9–13</td>
<td>1.17**</td>
<td>1.33**</td>
</tr>
<tr>
<td>Economic adversity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Received income assistance, 9–13</td>
<td>2.01**</td>
<td>2.34**</td>
</tr>
<tr>
<td>Less than average Grade 9 achievement</td>
<td>9.27**</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note: Statistically significant predictors with odds ratios differing most from zero are bolded. M = mother married at time of child’s birth; U = mother unmarried at time of child’s birth.

* N = 62,739; ** p < 0.001; * p < 0.0125; † p < 0.0125; ‡ p < 0.001.

† Model 1 includes “Same but Previous” predictor; Model 2 does not include “Same but Previous” predictor.
gene-environment correlation and by controlling for reciprocal causation. Because stronger research designs tend to show smaller relationships, bootstrap methods were used to determine whether significant differences existed between effects in the population and sibling models (see Supplementary Appendix B). When examining the “failure to graduate from high school” models, a significant decrease in effect was seen in only one predictor – “two children, second born” was significant in the population model but not in the sibling model. The “ADHD, 14–18” model showed no differences. Since only one effect size differed in the population and sibling models, results from using the population sample appear robust and not hindered by familial confounding, social selection, or misclassification for the education and ADHD variables.
Because fit statistics and odds ratios change relatively little between the two types of samples, multilevel population analyses (often drawn from administrative data and consecutive birth cohorts) might be able to replace more traditional sibling samples (typically based on survey data). The greater number of cases provided by administrative data is particularly helpful in assessing importance (and statistical significance) when using many predictors.

**Model validation and sensitivity testing**

To determine that our model is not over-fit, we used 10-fold cross validation. Differences in c-statistics of the cross-validated model and the original model were small and not statistically significant, indicating that over-fitting was not an issue. Bootstrapping was used to determine the standard errors associated with our estimates; confidence intervals from the original models were very similar to those generated from the bootstrapped models (see Supplementary Appendix C).

Various sensitivity analyses determined how different analytic decisions impacted the results. Sensitivity testing examined models that: 1) remove variables only available between ages 9 and 13; 2) use neighbourhood of the oldest child at birth; 3) measure neighbourhood SES using Material and Social Deprivation indices; 4) analyze a population subsample for whom household income was available (details in Supplementary Appendix D). Overall, our findings were minimally affected and very robust. Additional comparisons examined whether predictors differed for males and females (see Supplementary Appendix E). With insufficient numbers of male-only siblings and female-only siblings to permit multilevel analyses, logistic regressions were run. Some differences in predictors were seen between males and females; given the weaker study design, these results are not as robust.

**DISCUSSION**

The adverse childhood experiences important for each outcome differ. While family instability and economic adversity more strongly affect failing to graduate from high school, adverse health events in childhood and early adolescence have a greater impact on late adolescent ADHD.

Almost every study measure important in predicting high school graduation accords with the perspective put forward by Link and his collaborators. They postulate that individuals of relative wealth and without family problems are disproportionately able to take advantage of government programs – such as (high school) education. Earlier success (as indicated by Grade 9 achievement) predicts later success. Odds ratios show some general similarities across the different measures for predicting graduation, with younger children in large families and children in poor, unstable families at particular risk.

The predictors of ADHD in late adolescence are more difficult to categorize. The family-level analyses indicate the ADHD diagnosis to be somewhat related to access to urban physicians and to having an older mother. Several of the family structure history measures (parental divorce and two or more transitions) remain significant but mother’s unmarried status seems less important. This – and the lack of relationship between family receipt of income assistance and the ADHD diagnosis – suggests a smaller role for economic circumstances. Prior health conditions (particularly diagnoses pertaining to externalizing mental health conditions) generate consistently significant odds ratios. The very high odds ratio related to ADHD diagnoses at an earlier age highlight this condition’s continuity. Severe family problems significantly contribute in the population analysis; the relatively small number of siblings coded differently on this variable may contribute to the lesser statistical significance of these problems.

**CONCLUSION**

This study highlights the importance of including both social determinants of health and health conditions in childhood and early adolescence when examining cognitive and emotional outcomes. The large number of cases, which facilitated assessment of the predictors of adolescent intellectual and emotional development, is a major study strength. The longitudinal nature of this research allows making stronger inferences about the impact of childhood adversity on health and education; all of the covariates in the models precede the outcome. Population and sibling samples provide complementary approaches to evaluating the importance of predictors. Our approach should permit both developing valid risk prediction tools (currently used clinically for patient decision making) and identifying those who might benefit from specific interventions. Applying such tools at the population level can give insight into these outcomes for a province, region or nation and further quantify the burden of risk in the population.

Administrative data do have limitations. Symptoms may occur well before an official diagnosis and family problems often precede a formal change (such as divorce or remarriage). Diagnostic identification is less than perfect, and some variables will be lacking. But linkable databases can facilitate “expansion” through the creation of new variables, such as a measure combining residential mobility and school mobility. Over time, additional measures (such as child maltreatment) will become available. Incorporating both surveys and new government datasets (such as housing and justice) into information-rich environments may further increase both the explainable variance and the number of interesting outcomes.

Time-varying measures are critical to better model a life course perspective; however, the timing of such events is an important consideration. Examining several stages of childhood (for instance age groups 0–3, 4–8 and 9–13) would indicate which predictors appear most influential when, and provide important information regarding the appropriate timing of interventions. Maternal data (produced by linking mother’s and child’s files) should support both a family perspective and new modes of analysis. Unique research possibilities are being created by the availability of longitudinal linkable data.

**REFERENCES**

WHAT PREDICTORS MATTER


Received: June 5, 2015
Accepted: October 11, 2015

RÉSUMÉ


MÉTHODE : Les données administratives du Centre d’élaboration et d’évaluation de la politique des soins de santé, au Centre de la politique des soins de santé du Manitoba, ont servi à mener l’un des protocoles d’étude observationnelle les plus robustes: la modélisation multiniveau d’échantillons d’une grande population (n = 62 739) et de frères et sœurs (n = 29 444). Les variables incluses étaient les caractéristiques du quartier, des indicateurs de stabilité familiale, ainsi que les états de santé mentale et physique durant l’enfance et l’adolescence.


CONCLUSION : Ces résultats éclairent le rôle de plusieurs variables familiales et sociales et peuvent servir à créer des outils de prédiction fiables et valides pouvant identifier les personnes à haut risque. L’application d’un tel outil à l’échelle d’une population donnerait une idée du fardeau futur de ces effets dans une région ou un pays et permettrait de chiffrer davantage le fardeau du risque dans la population.

MOTS CLÉS : déficit de l’attention avec hyperactivité; niveau d’instruction; risque; études longitudinales