RESEARCH PAPER

Methods of the International Tobacco Control (ITC) Four Country Survey

M E Thompson, G T Fong, D Hammond, C Boudreau, P Driezen, A Hyland, R Borland, K M Cummings, G B Hastings, M Siahpush, A M Mackintosh, F L Laux

....

See end of article for authors' affiliations
....

Correspondence to:
Mary E Thompson, PhD, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada; methomps@uwaterloo.ca

This paper outlines the design features, data collection methods and analytic strategies of the International Tobacco Control (ITC) Four Country Survey, a prospective study of more than 2000 longitudinal respondents per country with yearly replenishments. This survey possesses unique features that sets it apart among surveys on tobacco use and cessation. One of these features is the use of theory-driven conceptual models. In this paper, however, the focus is on the two key statistical features of the survey: longitudinal and "quasi-experimental" designs. Although it is often possible to address the same scientific questions with a cross-sectional or a longitudinal study, the latter has the major advantage of being able to distinguish changes over time within individuals from differences among people at baseline (that is, differences between age and cohort effects). Furthermore, quasi-experiments, where countries not implementing a given new tobacco control policy act as the control group to which the country implementing such a policy will be compared, provide much stronger evidence than observational studies on the effects of national-level tobacco control policies. In summary, application of rigorous research methods enables this survey to be a rich data resource, not only to evaluate policies, but also to gain new insights into the natural history of smoking cessation, through longitudinal analyses of smoker behaviour.

The recently adopted Framework Convention on Tobacco Control (FCTC) has placed national-level tobacco control policies in the spotlight. Surprisingly, much of the policy research supporting the FCTC is based on secondary or cross-sectional analyses and/or ecological study designs.[1] These research designs are limited in their ability to shed light on the underlying causal mechanisms of policy impacts. Moreover, many of the published studies are based on single-country data, making it difficult to know if the same effects would be observed in other countries that might adopt the same policy. There have been virtually no international studies involving natural experiments (also known as quasi-experiments), in which one country that is implementing a tobacco control policy is compared to other countries in which no such policy has been implemented. The lack of a control group impairs the ability to measure the effects of a new policy.[1]

The objective of the International Tobacco Control (ITC) Policy Evaluation Project is to apply rigorous research methods to evaluate the psychosocial and behavioural effects of national-level tobacco control policies. The ITC Project uses multiple country controls, longitudinal designs, and theory-driven mediational models that allow tests of hypotheses about the anticipated effects of given policies.[1]

The ITC Policy Evaluation Project began in 2002 as a prospective cohort study tracking and comparing the impact of national level tobacco policies among representative samples of adult smokers in four countries: the United States, Canada, the United Kingdom, and Australia. In 2003, the collection of ITC Surveys was expanded to include smokers from Ireland and a new cohort of smokers from the UK, to evaluate the 2004 Ireland smoke-free policy.[2] In 2005, the collection of ITC Surveys was further expanded to include cohorts of smokers in Thailand and Malaysia.

This paper outlines the design features, data collection methods and analytic strategies utilised by the ITC Project. Although the focus of this paper is on the original four countries, table 1 outlines the survey designs and data collection dates of the nine countries currently involved in the ITC Project.

The ITC Four Country Survey (ITC-4) is a prospective cohort study of over 2000 adult smokers per country. These were recruited by probability sampling methods in each of the four countries in 2002. In subsequent follow-up surveys of the cohort, the recruited samples have been replenished after attrition to ensure a sample size of at least 2000 per country at each wave. This strategy also allows repeated cross-sectional analyses in each country. Further details of the sampling design used by the ITC-4 can be found in a technical report,[3] available on the ITC Project website at http://www.itcproject.org/.

DESIGN FEATURES

Supporting the analytic strategies of the ITC Project are two key design features that allow the use of quasi-experimental designs or "natural experiments".[4] The first is the international nature of the survey, in which one group (country) that is exposed to a policy (or to the highest strength of the policy) is compared to another group or other groups not exposed to the policy (or exposed to lower strength version). The second is the use of longitudinal samples, in which groups and their members are differentially exposed over time to policy changes. Longitudinal studies provide internal controls in the sense that it is possible to isolate changes over time within individuals from differences in baseline levels or cohort effects.[5] The other features of the ITC-4 Survey are described below.

....

Abbreviations: AAPOR, American Association for Public Opinion Research; CASRO, Council of American Survey Research Organizations; FCTC, Framework Convention on Tobacco Control; GEE, generalised estimating equations; ITC, International Tobacco Control; ITC-4, International Tobacco Control Four Country Survey; RDD, random-digit dialling; SES, socioeconomic status; SSI, Survey Sampling International

Table 1 Summary of the different International Tobacco Control (ITC) surveys

Country	Initial survey	Subsequent surveys (to date)	Survey design description
Australia	2002	2003, 2004, 2005	2000 adult smokers surveyed via RDD telephone interview in 2002 and yearly with replenishment through 2008
Canada	2002	2003, 2004, 2005	Same as Australia
UK	2002	2003, 2004, 2005	Same as Australia
USA	2002	2003, 2004, 2005	Same as Australia
Ireland	2003	2005, 2006	1000 adult smokers surveyed via RDD telephone interview in 2003, 2005 (with replenishment) and 2006 to access the response to the Ireland smoke-free law
Malaysia	2005	2006	2000 adult smokers, 1500 adult non-smokers, 1000 teenagers (aged 13 to 17 years) surveyed face-to-face via household canvass in 2005, with yearly replenishments in 2006 and 2007
Thailand	2005	2006	Same as Malaysia, but with no adult non-smokers recruited
South Korea	2005	Undetermined	1000 adult smokers surveyed via RDD telephone interview in 2005 with plans for repeat surveys pending available resources
China	2006	Undetermined	800 adult smokers and 200 non-smokers surveyed face-to face via household canvass in 2006 in each of 7 cities with plans for repeat surveys pending available resources

RDD, random-digit dialling

Timeframe

Data collection periods extended from October to December 2002 for Wave 1, from May to August 2003 for Wave 2 (a mean of 6.7 months after wave 1), and from June to December 2004 for Wave 3 (a mean of 12.8 months after wave 2). Beginning with Wave 4, these annual survey waves are or will be conducted from August to December, through 2009.

The timing of the waves was chosen to allow for the evaluation of a significant number of policies and trends. For example, the UK introduced more comprehensive warning labels and restrictions on advertising and promotion in early 2003, between Waves 1 and 2. The UK also introduced a ban on misleading product descriptors such as ''light'' and ''mild'' in September 2003.[6] Policy changes have also occurred in other ITC countries, such as in Canada, where more comprehensive restrictions on tobacco sponsorships and promotions were implemented between Waves 1 and 2.

Sampling frames and sampling at Wave 1

At the time of recruitment, ITC-4 participants are adult smokers—that is, individuals 18 years or older who have smoked more than 100 cigarettes in their lifetimes and who have smoked at least once in the past 30 days. The within country sample size of 2000 smokers was chosen to allow for the possibility of detecting changes in national proportions with high statistical power, even though the sample size crossing two waves is reduced by attrition. Power calculations were performed for several kinds of effects spanning the length of our conceptual model.[1 3]

At Wave 1, within each country, the population was stratified into several geographic regions. The number of strata assumed for weighting purposes ranged from 12 in the USA and the UK to 14 in Canada and Australia. Quotas were assigned for the numbers of respondents in each of the strata, in order to ensure representation proportional to a measure of regional population size. Eligible households were then randomly selected, using random-digit dialling (RDD) methods, until the within stratum quotas were met. A household was deemed to be eligible if it contained at least one eligible smoker. In households with multiple eligible smokers, the Next Birthday Method[7] was used to select a single respondent. No substitution within the household was permitted, except where it was known that the selected respondent would be absent for the entire fieldwork procedure. In three of the countries—Canada, the USA, and the UK—samples have been generated by Survey Sampling International (SSI) using their RDD B list assisted methodology.[8] In Australia, a comparable sampling frame was developed especially for the ITC-4.[3]

Smokers who have subsequently quit have been retained in the sample because their responses as quitters to many of the questions are of interest. Furthermore, retaining former smokers allows observation of transitions to other types of tobacco products, relapse back to smoking, and subsequent efforts to stop smoking among those who do relapse.

Cohort replenishment

In order to ensure that the number of completed surveys at each wave is at least 2000 per country, respondents lost to attrition have been replaced. Replenishments have been carried out using the same sampling design and calling protocol as in Wave 1 recruitment. Any given replenishment sample is thus representative of the population at the time of data collection for the new wave to which they correspond, rather than those lost to follow-up. As the survey proceeds, data from continuing and replenishing respondents can be examined and compared to assess the influence of ''time-in-sample'' on the outcome variables. For example, it is possible to examine the extent to which the replenishing respondents behave differently from their counterparts recruited at earlier waves, and the extent to which continuing respondents are still a representative sample of the population of smokers. Panel attrition at each wave can be modelled as depending on age, sex, education and other factors, including health status and smoking cessation intentions at previous wave(s).

DATA COLLECTION METHODS
Recruitment into the Wave 1 cohort

A 10-minute recruitment survey was first conducted to screen for eligibility and ascertain consent. Respondents were asked if they were willing to participate in a 35-minute survey on smoking being conducted by an international group of universities and research institutions in four countries. They were also told they would receive a small payment to thank them for their time, and that they would be contacted in approximately six months to complete a second survey (for which they would receive a second payment). Furthermore, they were assured that their responses would be kept strictly confidential. Those who consented and were eligible were phoned about one week later for completion of the main survey. Consent was obtained separately for each wave.

Compensation

A thank you letter and financial compensation were mailed immediately after the recruitment call, so as to arrive before participants were called back to complete the main survey. Compensation amounts were roughly standardised across countries: $10 US, $15 CDN, $15 AUS, or a £7 voucher for Boots (a retail pharmacy and store). Over 90% of respondents

in each of the four countries had received this compensation by the time of the main survey call. This kind of pre-survey compensation has been shown in randomised experiments on incentives to increase response rates, sometimes by over 10 percentage points.[9–11]

For subsequent waves, prior respondents were sent a letter with the compensation, approximately one week before being re-contacted. In that call, respondents were given the option of completing the interview immediately, or scheduling it at a later date. Recruitment of new respondents proceeded in the same manner as for Wave 1.

Calling protocol

Several steps were undertaken to minimise bias and maximise response rates. First, to avoid call-scheduling bias, recruitment calls were conducted at various times of the day and on different days of the week. Second, once a respondent had completed a survey, strong efforts were made to follow him/her up for the next interviews. This included collecting alternate contact information to aid in following respondents who had moved between waves. In the event that a respondent did not keep a main survey appointment, up to 25 attempts to follow-up were made at varying times of day. In addition, respondents could complete the main survey during two or more calls if requested. Moderate attempts were made to convert refusals, primarily through emphasising the importance of the research.

Waves 1 and 2 of the survey were conducted in Canada and the USA by Environics Research Group, and in Australia and the UK by Roy Morgan Research. Waves 3 and 4 were conducted in all four countries by Roy Morgan Research. Senior representatives from each firm participated in the protocol design, in order to ensure standardisation of survey administration and calling protocol across both firms. All calling specifications, final questionnaires, and daily reports were reviewed and monitored by the ITC Research Team at the University of Waterloo to maintain consistency across survey firms and countries.

Questionnaire development

The ITC-4 questionnaire was developed by our international team of experts on tobacco control, with varied backgrounds including psychology, public health, economics, community medicine, marketing, sociology and statistics/biostatistics. A pilot survey was conducted just before Wave 1 to test the questionnaire and the study protocol, as well as further refining the survey measures. For this pilot survey, approximately 125 participants in each country completed both the recruitment and the main surveys. Modifications to wording and question framing were made as a result of this pre-testing. The questionnaire has been revised at each subsequent wave, but the core of the instrument has remained essentially the same to facilitate comparisons and modelling over time.

Apart from minor variations in wording to account for national differences in colloquial speech (for example, bar/pub), the same questionnaire was used in all four countries for English speaking respondents. A French translation of the questionnaire was administered to French Canadian respondents, by interviewers fluent in the language.

ITC-4 SAMPLE
Wave 1 recruitment

Survey lengths and sample sizes at recruitment for each of the four countries are provided in table 2. All call attempts were assigned an American Association for Public Opinion Research (AAPOR) disposition code.[12] For each country, outcome rates were calculated according to AAPOR and to the Council of American Survey Research Organizations

(CASRO).[3][13] Table 2 shows the cooperation rate (that is, the proportion of those screened who agreed to complete the recruitment survey) and the AAPOR Response Rate #4 for each country. Compared to the cooperation rate, this latter measure of the response rate accounts for the fact that numerous individuals could not be contacted or screened for eligibility. Furthermore, it has the distinguishing feature that the proportion of cases of unknown eligibility that are actually eligible is estimated rather than assumed to be one. With smoking rates between 20–30% in each of the four countries, only a fifth to a third of the adult population were potentially eligible for the survey. Hence, estimating the number of eligible cases is of great importance in calculating meaningful response rates. This was done using country specific eligibility rates.[3]

Despite high cooperation rates in all countries, the country response rates, at least in North America, are lower than those typically reported in the literature. It should be noted that rates reported in the literature rarely adhere to AAPOR standards. The ITC-4 rates are based on these conservative standards and should be interpreted as such. An examination was carried out of response rates of the largest recent telephone surveys on smoking or health behaviour, one in the USA and two in each of the other three countries. Wherever possible, the AAPOR response rate equation was applied to disposition data, to assist in comparing response rates with those of the ITC-4. The response rates for the comparison surveys were 70% (both surveys) in Canada, 55% in the USA, 10% and 56% in the UK, and 46% and 45% in Australia. Given these figures, our response rates in the UK and, particularly, in Australia are in good agreement with the ones from recent telephone surveys in these countries. Although our response rate is lower in Canada, both of the surveys mentioned above were conducted on behalf of the government of Canada, and greatly benefited from their official status. The ITC-4 response rate for the USA sample is considerably lower than for the other countries. This is mainly due to the relatively high number of terminations before completion of the screening process and to a trend towards lower response rates for RDD surveys in North America.[14] For example, median response rates for longitudinal Centers for Disease Control and Prevention phone surveys declined from over 71% in 1993 to below 49% in 2000.[15] Anecdotal evidence suggests that response rates from more recent surveys, yet to be published, have decreased even further. The ITC-4 response rate may be further evidence of this trend.

Wave 1 main survey

The overall follow-up or retention rates from Wave 1 recruitment to main surveys are given in table 3, and vary from 85.5% in the USA to 89.8% in Australia. An analysis of loss to follow-up (or attrition) and retention by respondent categories was carried out. As expected, retention rates were lower for males in every broad age group (except for 50+ in the UK), for respondents in large urban areas, and for non-white people in Canada and the USA.[16] Furthermore, retention rates were somewhat lower in the 18–34 age group for both sexes. The dependence of attrition and retention on educational attainment was minor and not consistent across

| Table 2 | Recruitment survey rates: Wave 1 |

	Canada	USA	UK	Australia
Number recruited	2507	2500	2730	2566
Cooperation rate	82.3%	83.2%	78.7%	78.8%
Response Rate AAPOR#4	49.5%	25.6%	37.8%	45.8%
Survey length (min)	11.8	13.2	10.3	9.1

Table 3 Wave 1 main survey rates

	Canada	USA	UK	Australia
No. from recruitment survey	2507	2500	2730	2566
Refused	4.5%	4.7%	6.5%	3.7%
Non-contact	7.3%	9.8%	5.6%	6.5%
Follow-up rate	88.2%	85.5%	87.9%	89.8%
No. of completed interviews	2214	2138	2401	2305
Survey length (min)	43.4	44.6	38.6	38.7

the countries. Attrition was higher for those who rated their health as ''poor'' in Canada and Australia, but neither in the USA nor the UK. On the basis of these results it was decided not to incorporate education or perceived health in the adjustments for the main survey weights (see Survey weights).

It is usually desirable to account, at least partially, for the determinants of retention and attrition when analysing the survey results. Thus, statistical models for ITC-4 data should include age, sex and white versus non-white in the USA. Other predictors of attrition that could affect some analyses are marital status (single respondents are more often lost to follow-up) and quit attempt history.

Representativeness of the sample
Survey weights are used to account and adjust for uneven representation. To assess the representativity of the ITC-4 sample, the demographic profiles of the unweighted ITC-4 samples have been compared with the profiles of the national benchmark surveys used for calculating the weights.[3 16] Differences in the demographic profiles between the ITC-4 survey and benchmarks are modest. The Wave 1 sex proportions from the unweighted ITC-4 estimates and the national benchmarks are very close for Canada and the USA, and within 5 percentage points for Australia and the UK. The age and race/ethnicity profiles are also very close. The ITC-4 estimates of educational profiles are in agreement with national benchmarks in Canada, the USA and the UK. (The Australian benchmark did not measure education in the same way as did the ITC-4.)

The ITC-4 weights are constructed so that weighted estimates of the number of smokers in any given age/sex/region group combination agree with the corresponding national benchmarks, thus eliminating non-response biases with respect to those three characteristics. For other variables, if non-contact and non-response occur randomly (see Non-response and attrition), no bias is introduced and the validity of the estimates is unaffected. In other words, low response rates lead to biased estimates only to the extent that non-respondents differ from respondents on the characteristics of interest. Frequently, large differences in response rates have shown only minor effects on key estimates.[15 17 18] Indeed, response rate differences as large as 45% have been found to yield a predicted difference in smoking prevalence estimates of only 1.5 percentage points.[15]

To verify that non-response was approximately random, key measures of smoking status from the ITC-4 were compared with results from the national benchmark surveys of adult smokers. The measures were cigarettes per day, quit attempts in the last 12 months, whether seriously considering or planning to quit in the next six months, and whether having first cigarette of the day within five minutes of waking. ITC-4 Wave 1 data for these key measures are in good agreement with those of the benchmark surveys, with the exception of slightly higher measures of daily consumption and addiction in the ITC-4 sample.[3]

Considering that non-response has been associated with lower socioeconomic status (SES)[19] and given the relationship between tobacco use and SES,[20] the ITC-4 data might be expected to yield a sample of lighter, less-addicted smokers than surveys with higher response rates, such as the national benchmarks in Canada and the USA. However, the data suggest that the ITC-4 has been relatively successful in surveying traditionally hard-to-reach respondents. Furthermore, the differences between ITC-4 and national benchmarks are modest and often no greater than the differences among national benchmarks themselves.

Nevertheless, some variables of interest may well be associated with response behaviour, even after controlling for the variables used to construct the sampling weights. In such cases, non-contact and non-response are non-random, and must be modelled using statistical procedures that require additional assumptions (see Non-response and attrition).

Retention and replenishment, Waves 2 and 3
The number of respondents retained, as well as the corresponding retention rates, from the first two cohorts of the ITC-4 (sampled in 2002 and 2003, respectively) at Waves 2 and 3 are given in table 4. In addition to these numbers, table 4 gives the sample sizes of the first three ITC-4 cohorts.

At Wave 2, the proportions of Wave 1 respondents retained were respectively 75.8%, 62.8%, 77.7%, and 81.4% in Canada, the USA, the UK and Australia. A new sample was recruited using the Wave 1 design in each country, with sizes 517 in Canada, 684 in the USA, 255 in the UK and 258 in Australia (table 4).

Retention rates were also calculated according to smoking status and quitting intentions in Wave 1. In Canada and the USA, retention rates were considerably lower among those who had said in Wave 1 that they had quit by the time of the main survey (56.0% in Canada and 36.1% in the USA). The sample sizes for this group are small, but the phenomenon bears watching. Retention rates depended very little on quitting intentions.

At Wave 3, the proportions of Wave 2 respondents retained were respectively 70.0%, 58.3%, 65.1%, and 75.2% in Canada, the USA, the UK and Australia. The same replenishment strategy was used as in Wave 2, with sizes 543 in Canada, 889 in the USA, 586 in the UK and 532 in Australia (table 4).

The numbers retained from Wave 1 to Wave 3 were 1201 in Canada, 799 in the USA, 1328 in the UK and 1377 in Australia (table 4). The difficulty in retaining respondents in the USA is reflected in the relatively large number of cases where the respondent could not been traced, despite use of the alternate contact information obtained at recruitment.

STATISTICAL METHODS
Survey weights
The sampling design was chosen to provide a random, unbiased and representative sample of adult smokers within each geographic stratum. However, as with most surveys, the ITC-4 sample is subject to some disproportionate selection and under-coverage of population subgroups. In order to adjust for disproportionate selection of adult smokers in subgroups, weights have been calculated for each respondent. The following describes the procedures for calculating these weights.

Recruitment and Wave 1 sampling weights
First, each recruited respondent was assigned a factor to account for the number of residential phone lines and the number of adult smokers in the household. The result was then multiplied by another factor to produce another preliminary weight for each respondent. The sum of these

Country	Cohort	Wave 1	Wave 2		Wave 3	
		n	n	Retained	n	Retained
Canada	Wave 1	2214	1679	75.8%	1201	54.2%
	Wave 2	NA	517	NA	362	70.0%
	Wave 3	NA	NA	NA	543	NA
	Total	2214	2196	NA	2106	NA
USA	Wave 1	2138	1342	62.8%	799	37.4%
	Wave 2	NA	684	NA	399	58.3%
	Wave 3	NA	NA	NA	889	NA
	Total	2138	2026	NA	2087	NA
UK	Wave 1	2401	1865	77.7%	1328	55.3%
	Wave 2	NA	255	NA	166	65.1%
	Wave 3	NA	NA	NA	586	NA
	Total	2401	2120	NA	2080	NA
Australia	Wave 1	2305	1876	81.4%	1377	59.7%
	Wave 2	NA	258	NA	194	75.2%
	Wave 3	NA	NA	NA	532	NA
	Total	2305	2134	NA	2103	NA

Table 4 Retention and replenishment at waves 2 and 3

NA, not applicable.

preliminary weights over respondents in any given stratum was proportional to the quota for that stratum. This compensates for differential achieved sampling fractions among strata. Finally, these preliminary weights were adjusted to produce recruitment weights w_r so that estimates of total numbers of smokers in age-sex groups (and white/non-white groups in the USA) agreed with current smoking prevalence numbers.[3]

From the calibrated recruitment weights w_r, an additional set of weights w_1 was created for the subset of respondents who completed the main survey entirely or partially (n = 9058). These w_1 weights were adjusted for attrition between the recruitment and the main survey, and then recalibrated. In this process, recruited respondents who did not complete the main survey, either entirely or partially, were assigned a weight of 0. Note that a respondent for whom it was impossible to determine smoking status was considered not to have completed the main survey.

Weights for Wave 2 and Wave 3

For Wave 2, the following sets of weights have been constructed:

1. Longitudinal weights for longitudinal or cohort analyses based on respondents who completed both Waves 1 and 2 main surveys; these are the weights w_1 adjusted for attrition and re-calibrated to the same prevalence numbers.
2. Cross-sectional weights for analyses involving new Wave 2 respondents only; these are calibrated to prevalence numbers at the time of Wave 2. Such analyses are relevant to the investigation of whether new respondents exhibit different behaviour than continuing respondents.
3. Cross-sectional weights for analyses involving all Wave 2 respondents; like the other cross-sectional weights above, these weights are calibrated to assumed prevalence numbers at the time of Wave 2.

For Wave 3, analogous sets of weights have been constructed.

Note that, at each wave, the longitudinal sample is a little less representative of its original population because of attrition. Consequently, the weights become slightly more variable. However, the coefficients of variation of the weights at Waves 1, 2 and 3 remain reasonable at around 0.46, 0.53 and 0.63, respectively.

Respondents to the ITC-4 are selected according to a complex survey design, involving stratification and unequal probabilities of selection. Consequently, standard error estimators need to be adjusted to take the "design effect" of the survey and the calibration of the weights into account. Formulae for such standard error estimators are readily available from sampling theory.[21] Some of these formulae are built into statistical packages like SUDAAN,[22] WesVar,[23] Stata[24] and SAS[25] (PROC SURVEYMEANS and PROC SURVEYFREQ). Alternatively, and as is often done in practice, the results of these formulae can be approximated by bootstrapping techniques that respect the sampling design and weight construction.[26 27] Sets of bootstrap weights have been computed for the ITC-4 to this end.

Cross-sectional and longitudinal analyses

The international and longitudinal study design of the ITC Surveys naturally leads to the following core analytical strategies:

1. Cross-sectional comparisons between countries or jurisdictions with policy differences.[28]
2. Cross-sectional analyses of the relationships between proximal (that is, policy specific) variables and health-related outcomes (for example, cessation intentions and recent quitting attempts), within or between jurisdictions or policy environments.
3. Longitudinal comparisons of changes in proximal and outcome variables between jurisdictions where relevant policy changes have occurred and where they have not.[2 29]
4. Studies of the evolution of proximal and outcome variables over time.
5. Longitudinal studies of how proximal variables predict outcomes in the subsequent waves, and how baseline differences in policies affect those relationships.[30]
6. Transition analyses, in which outcomes are modelled as dependent on their values in earlier wave or waves.
7. Latent variable modelling of longitudinal trajectories, to measure the extent to which smoking behaviour and quitting behaviour are predicted by the same or different psychosocial factors.[31 32]

Multi-year data may be regarded in several ways. First, the data provide a series of repeat cross-sectional samples allowing tests of hypotheses and construction of models that are part of analytical strategies 1 and 2. Moreover, by allowing comparisons between two or more waves, the data structure enables testing hypotheses that can be expressed in

terms of analytical strategy 3. The large overlap between yearly samples implies that observations are not and can not be treated as independent; correlation between measurements from the same respondent must be taken into account. To this end, generalised estimating equations (GEE) are widely used.[33]

Second, part of the sample is longitudinal and provides repeated measures data from the same respondents over several waves, enabling statistical modelling and testing of scientific questions that can be formulated in terms of analytical strategies 4 and 5. As previously mentioned, longitudinal studies have a substantial advantage over repeated cross-sectional studies, since longitudinal respondents act as their own control, allowing the differentiation of changes over time within individuals from differences in baseline levels. Various statistical methods can be used in studying longitudinal data; these include: repeated measures ANOVA and MANOVA, linear mixed models or mixed effects models, and generalised linear mixed models. In the last three types of models, respondents' slopes and intercepts, regarded as random, may be modelled as dependent on demographic, personal and environmental context variables (see Skrondal and Rabe-Hesketh[34] for a general treatment of random effects, often referred to as latent variables).

In longitudinal surveys, care must be taken in the selection of survey weights, particularly when more than one cohort is present as is the case in the ITC-4 with yearly replenishments.[35] It should also be kept in mind that the longitudinal sample and the finite population may evolve in different ways as time goes on.[36]

Mediation models

As noted by Fong et al,[1] it is of interest to test whether the effects of policies on downstream distal variables (*psychosocial mediators*) and behavioural end points (for example, quit attempts) are *mediated* by the proximal variables (*policy-specific variables*). Analytic methods for conducting such mediational analyses are well-known in the psychosocial literature.[37] Because the general analytic framework of the ITC-4 is more complex—employing generalised linear mixed models and complex survey designs—methods for generalising these well-known approaches into our framework will require some development and testing.

Non-response and attrition

The ITC-4 longitudinal survey weights are designed and constructed to correct for some of the biases due to attrition from one wave to the next. Specifically, analyses containing only variables that are "missing at random" (that is, for which the probability of missing does not depend on the dependent variable)[38] will be unbiased. In refined analyses, attrition may be modelled in terms of such variables as sex, age, education, health and marital status. Where attrition can be explained by these variables, the cohort weights may be adjusted by dividing by scores for the "propensity to respond" derived from the attrition model.[39] These adjusted weights can then be used in GEE, mixed-effects and other models mentioned in the "Cross-sectional and longitudinal analyses" section.

CONCLUSION

The ITC-4 Survey is the first survey to use an international prospective cohort design to evaluate the impact of tobacco control policy measures as they are introduced in participating countries. Innovative design features and the experience gained through the development of carefully specified and closely monitored protocols will make the use of the ITC methodology possible on a wider scale, as more countries join the project.

Methodological challenges include accounting for the varying "telephone" culture from country to country, and designing for differential attrition among age groups. The relative simplicity (among complex designs) of the basic stratified random sampling design makes it possible to use survey weights to remove some of the biases due to non-response and attrition, without having a large impact on standard errors.

The survey provides a very rich data resource, not only to evaluate policies and the mediators of their impact, but also to gain new insights into the natural history of smoking and cessation, through longitudinal analyses of smoker behaviour.

. . . .

Authors' affiliations

M E Thompson, G T Fong, D Hammond, C Boudreau, P Driezen, University of Waterloo, Waterloo, Ontario, Canada
A Hyland, K M Cummings, Roswell Park Cancer Institute, Buffalo, New York, USA
R Borland, M Siahpush, The Cancer Council Victoria, Carlton, Victoria, Australia
G B Hastings, A M Mackintosh, University of Stirling, Stirling, UK
F L Laux, Northeastern State University, Tahlequah, Oklahoma, USA

Competing interests: no competing interests.

Ethics clearance: This study has been reviewed and received ethics clearance from the Institutional Review Board or Research Ethics Board at each of the following institutions: Roswell Park Cancer Institute, the University of Waterloo, the University of Illinois at Chicago, The Cancer Council Victoria, University of Strathclyde, University of Stirling and The Open University.

REFERENCES

1 **Fong GT**, Cummings KM, Borland R, *et al.* The conceptual framework of the International Tobacco Control (ITC) Policy Evaluation Project. *Tob Control* 2006;**15**(suppl III):iii3–11.
2 **Fong GT**, Hyland A, Borland R, *et al.* Reductions in tobacco smoke pollution and increases in support for smoke-free public places following the implementation of comprehensive smoke-free workplace legislation in the Republic of Ireland: findings from the ITC Ireland/UK Survey. *Tob Control* 2006;**15**(suppl III):iii51–8.
3 **Hammond D**, Fong GT, Thompson ME, *et al.* International Tobacco Control Policy Evaluation [ITC Four Country Survey] wave 1 technical report. 2004 http://www.itcproject.org.
4 **Cook TD**, Campbell DT. *Quasi-experimentation: design and analysis issues for field settings.* Boston: Houghton Mifflin, 1979.
5 **Diggle PJ**, Liang K-L, Zeger SL. *Analysis of longitudinal data,* 2nd ed. Oxford: Oxford Science Publications, 2002.
6 **ASHUK**. ASH briefing: The UK ban on tobacco advertising. 2005 http://www.ash.org.uk/html/advspo/html/timetable.html (Accessed 21 Feb 2005).
7 **Survey Sampling International**. *RDD telephone random digit samples,* http://www.surveysampling.com/.
8 **Binson D**, Canchola JA, Catania JA. Random selection in a national telephone survey: a comparison of the Kish, next-birthday, and last-birthday methods. *Journal of Official Statistics* 2000;**16**:53–60.
9 **Singer E**, van Hoewyk J, Maher MP. Experiments with incentives in telephone surveys. *Public Opinion Quarterly* 2000;**64**:171–88.
10 **Lynn P**, Taylor B, Brook L. Incentives, information and number of contacts: testing the effects of these factors on response to a panel survey. *Survey Methods Centre Newsletter* 17. National Centre for Social Research, London, UK, 1997.
11 **Martin E**, Abreu D, Winters F. Money and motive: effects of incentives on panel attrition in the survey of income and program participation. *Journal of Official Statistics* 2001;**17**:267–84.

12 **American Association for Public Opinion Research**. *Standard definitions: final dispositions of case codes and outcome rates for surveys*. Lenexa, Kansas: AAPOR, 2000.

13 **Council of American Survey Research Organizations**. *On the definition of response rates: a special report of the CASRO Task Force on completion rates*. Port Jefferson, New York: CASRO, 1982.

14 **Flynn CA**. Biener L, Roman AM. Consequences of declining response rates for estimating population smoking prevalence. National Conference on Tobacco or Health. San Francisco, 2002.

15 **Wortley P**, Mariolis P. Impact of response rates on estimates of smoking prevalence: comparison of state-specific estimates from the BRFSS and the CPS Tobacco Use Supplement. American Association for Public Opinion Research (AAPOR) Conference, March, 2002.

16 **Anon**. Supplementary tables, the International Tobacco Control Policy Evaluation Project, Waterloo, ON: http://www.itcproject.org.

17 **Keeter S**, Miller C, Kohut A, *et al*. Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly* 2000;**64**:125–48.

18 **Curtin R**, Presser S, Singer E. The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly* 2000;**64**:413–28.

19 **Koch A**, Prost R. Nonresponse in Survey Research. *Proceedings of the Eighth International Workshop on Household Survey Nonresponse*. Mannheim: ZUMA, 1998.

20 **US Department of Health and Human Services**. Reducing tobacco use: a report of the Surgeon General. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Atlanta, Georgia, 2000.

21 **Thompson ME**. *Theory of sample surveys*. London: Chapman and Hall, 1997.

22 **Research Triangle Institute**. *SUDAAN*, Research Triangle Park, North Carolins; RTI.http://www.rti.org/sudaan/.

23 **Westat**. *WesVar*. Rockville, Maryland: Westat, http://www.westat.com/wesvar/.

24 **StataCorp LP**. *Stata*. College Station, Texas: StataCorp LP, http://www.stat.com/.

25 **SAS Institute, Inc**. *SAS*. Cary, North Carolina: SAS Institute, http://www.sas.com/.

26 **Efron B**, Tibshirani RJ. *An introduction to the bootstrap*. New York: Chapman and Hall, 1993.

27 **Lohr SL**. *Sampling: design and analysis*. Pacific Grove, California: Duxbury Press, 1999.

28 **Borland R**, Yong H-H, Siahpush M, *et al*. Support for and reported compliance with smoke-free restaurants and bars by smokers in four countries: findings from the International Tobacco Control (ITC) Four Country Survey. *Tob Control* 2006;**15**(suppl III):iii34–41.

29 **Harris F**, Mackintosh AM, Anderson S, *et al*. Effects of the 2003 advertising/promotion ban in the United Kingdom on awareness of tobacco marketing: findings from the International Tobacco Control (ITC) Four Country Survey. *Tob Control* 2006;**15**(suppl III):iii26–33.

30 **Borland R**, Yong H-H, Cummings KM, *et al*. Determinants and consequences of smoke-free homes: findings from the International Tobacco Control (ITC) Four Country Survey. *Tob Control* 2006;**15**(suppl III):iii42–50.

31 **Muthén B**, Curran PJ. General longitudinal modeling of individual differences in experimental designs: a latent variable framework for analysis and power estimation. *Psychological Methods* 1997;**2**:371–402.

32 **Raudenbush SW**. Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annu Rev Psychol* 2001;**52**:501–25.

33 **Liang K-Y**, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;**73**:13–22.

34 **Skrondal A**, Rabe-Hesketh S. *Generalized latent variable modeling: multilevel, longitudinal and structural equation models*. Boca Raton: Chapman Hall/CRC, 2004.

35 **Kalton G**, Brick, JM. Weighting schemes for household surveys. *Survey Methodology* 1995;**21**:33–44.

36 **Goldstein H**. *The design and analysis of longitudinal studies*. London: Academic Press, 1979.

37 **MacKinnon DP**, Lockwood CM, Hoffman JM, *et al*. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods* 2002;**7**:83–104.

38 **Little RJA**, Rubin D. *Statistical analysis with missing data*. New York: Wiley, 1987.

39 **Rosenbaum PR**. *Observational studies*, 2nd ed. New York: Springer Verlag, 2002.