Calibrating Questionnaires with Weekly Diaries : An Application in Religious Behavior, Netherlands 1975 to 2005

Published Version:
Ettore Scappini (2021). Calibrating Questionnaires with Weekly Diaries : An Application in Religious Behavior, Netherlands 1975 to 2005. SOCIOLOGICAL METHODOLOGY, 51(1), 166-187 [10.1177/0081175020927438].

Availability:
This version is available at: https://hdl.handle.net/11585/761641 since: 2021-01-26

Published:
DOI: http://doi.org/10.1177/0081175020927438

(Article begins on next page)

Sociological Methodology

Calibrating questionnaires with weekly diaries: an application in religious behavior (Netherlands 1975-2005)

Journal:	*Sociological Methodology*
Manuscript ID	SMX-19-0006.R2
Manuscript Type:	Original Articles
Keywords:	Time use, Measurement problems, Calibrating questionnaires, Measured presence, Diaries

Abstract:	This paper presents an innovative approach to improve the power of questionnaires by combining them with weekly diaries. The aim is to show how it is possible to calibrate information collected from the questionnaire, which provides a distribution that is in general biased, with diary data, which are more accurate but cannot provide any distribution across a range of frequencies. These problems become even more pronounced when the object of analysis is a specific issue such as religious practice, the focus of this study. The suggested user-friendly model uses the more accurate diary data to adjust the distribution produced by the standard questions and enables researchers to obviate the problems of the two data collection methods. To present a practical application we will use the Time-budget surveys (TBO), conducted at five-year intervals between 1975 and 2005 in the Netherlands.

Title

Calibrating Questionnaires with Weekly Diaries: An Application in Religious Behavior; Netherlands 1975 to 2005

Author's name

Ettore Scappini

Ettore Scappini, PhD, is a Senior Lecturer in Sociology at the University of Bologna, Italy. His primary research interests are in the field of methodological problems in the social sciences with specific topics including measurement errors in social surveys with the use of indicators derived from stylized questionnaires and diaries, the validity of indicators for measuring religious behavior, and the problems of estimated APC models with specific applications in political behavior.

Affiliation

Department of Education Studies, University of Bologna, via Filippo Re, 6, 40129 Bologna Italy.

Keywords time-use; diaries; measured presence; calibrating questionnaires; measurement problems.

Corresponding author detail:

Email, ettore.scappini@unibo.it;

ORCID ID, https://orcid.org/0000-0002-4909-8737

Acknowledgments

The author would like to thank the anonymous reviewers for their helpful comments, which made an important contribution to clarifying the issues discussed. I am also grateful to Anne Roeters for some personal recommendations about the use of the TBO data.

Abstract

This article presents an innovative approach to improve the power of questionnaires by combining them with weekly diaries. The aim is to show how one can calibrate information collected from questionnaires, which provide a distribution that is in general biased, with diary data, which are more accurate but cannot provide a distribution across a range of frequencies. These problems become even more pronounced when the object of analysis is a specific issue, such as religious practice, the focus of this study. The suggested user-friendly model uses the more accurate diary data to adjust the distribution produced by the standard questions and enables researchers to obviate the problems of the two data collection methods. To present a practical application, we use the time-budget surveys (TBO) conducted at five-year intervals between 1975 and 2005 in the Netherlands.

1. INTRODUCTION

This article presents a method to improve the information provided by questionnaires by combining them with weekly diaries and applies it to the issue of religious behavior in the Netherlands. The aim is to show how one can calibrate information collected from questionnaires, which is generally more complete but subject to bias, with data collected from diaries, which are usually much more accurate but less detailed. These problems become even more pronounced when the object of analysis is a specific issue such as religious practice, the focus of this study.

Current research recognizes two main data collection tools for measuring the frequency with which an activity is carried out: the stylized questionnaire (hereafter, simply questionnaire) and the diary. In the former case, respondents are generally asked to use a pre-coded scale to indicate how frequently they perform a given activity within an established period of time, which could be a week, a month, or more commonly a year. In the latter case, respondents compile a daily or weekly diary in which they indicate which activity or activities they carry out at established intervals, typically 10 to 15 minutes. Both data collection methods have their advantages and drawbacks.

Questionnaires can be used to carry out extremely detailed surveys regarding the distribution of the intensity with which a given activity is carried out by a population, but the data collected are somewhat imprecise (Harms et al. 2019; Jacobs 1998; Juster, Ono, and Stafford 2003; Kan and Pudney 2008; Robinson 1985). It is difficult to recall accurately at a later date which activity was carried out or how frequently it was performed (Gershuny 2012). Furthermore, questionnaires are disproportionately prone to social desirability or demonstration effects, inasmuch as indicating that a given activity was carried out, or a failure to do so, implies, at most, passive action (Gershuny 2012).

Questionnaires thus allow researchers to obtain detailed distributions of the degree of intensity with which an activity is carried out in a given population over a relatively broad pre-established period, such as the preceding year. However, as this result may be subject to bias, its level of reliability can become problematic and can cause researchers serious difficulties when analyzing the relevant phenomenon. The problem of questionnaires' low precision level is exacerbated when a study addresses specific issues, such as those related to some form of obligation or expectation, also in moral terms, regarding anticipated behavior, such as in the case of religious behavior (Hadaway and Marler 2005; Hadaway, Marler, and Chaves 1993, 1998; Presser and Chaves 2007; Presser and Stinson 1998; Rossi and Scappini 2012).

Diaries, in contrast, can be used to carry out more precise studies on the average intensity with which a given activity is performed. The chronological structure of a diary makes it easier to record the timing and recollection of events (Belli 1998; Belli and Callegaro 2009). Any bias due to memory gaps during the compilation phase is relatively limited and drops even further when there are special events, typically of a less frequent nature (Al Baghal et al. 2014). Yet diaries could have the problem of "reactance," that is, a change in participants' behavior as an effect of their involvement in the study (Bolger, Davis, and Rafaeli 2003). In other words, when a diary needs to be filled out, the respondent is reminded of the behavior that needs to be reported, which may lead to a change in behavior. This phenomenon has been observed in specific limited contexts (Korotitsch and Nelson-Gray 1999), such as cases of drug abusers (Hay, Hay, and Angle. 1977), but there is little evidence that this effect poses a general threat to the validity of diaries (Barta, Tennen, and Litt 2012; Shiffman, Stone, and Hufford 2008). Finally, unlike with questionnaires, any falsification requires episodes to be actively invented. Consequently, as it is easier for respondents to behave honestly, there are fewer desirability effects (Gershuny 2012).

However, diary data are much less detailed than questionnaires, as almost nothing is known about the distribution of the intensity with which the given activity is carried out in the period when the diary is not compiled (Frazis and Stewart 2012; Scappini 2010). Therefore, although diaries make it possible to obtain a reliable measure of the average intensity with which an activity is performed in a given population, they cannot generate even a simple frequency distribution in reference to longer periods than the duration of the diary. This situation can present researchers with serious problems, as analyses often require one to identify groups of subjects that carry out a given activity with different degrees of intensity for relatively long periods. For example, studies on religious practice frequently seek to identify the group that attended a service at least once a month over the preceding year.

To overcome the difficulties arising from these two data collection methods, scholars have suggested finding a model that can calibrate the values obtained from questionnaires with those from diaries in such a way as to obtain data that are as complete and reliable as possible. This requirement led to many attempts—above all by Gershuny—to make combined use of the qualities of the two data collection tools. Kan and Gershuny (2006, 2009) showed that it is possible to calibrate questionnaires by combining two datasets: one derived from a survey that collected questionnaire and diary data from the same respondents, and the other from a questionnaire-based survey. This method was then perfected using the latest matching techniques (Borra, Sevilla, and Gershuny 2013; Walthery and Gershuny 2019). Following a similar path, Gershuny (2012) showed that by taking Bourdieu's concept of *habitus* as a starting point, one can use logistic regressions to estimate the probability of participating at least once over 14- or 28-day periods.

The disadvantage of these methods is the use of unusual theoretical concepts and relatively complex regression techniques. In contrast, we present a model that does not require any particular assumptions and is easier to apply.

To summarize, diary data are more accurate than questionnaire data, but they cannot generate a distribution across a range of attendance-frequency response categories. The standard question, on the other hand, provides a distribution, but it is biased upward because people generally tend to overstate their attendance frequency. The calibration proposed here uses the more accurate diary data to adjust the distribution produced by the standard question.

To present a practical application, we use time-budget surveys conducted at five-year intervals between 1975 and 2005 in the Netherlands.[1] The data collection methods used were a questionnaire, including items about religion, and a weekly diary.

This study takes shape over three sections. In the first of these, we present the data and discuss the different characteristics of the indicators, while in the second we develop the reasoning that led us to create the model and discuss some necessary premises for its application. The third section features comparisons highlighting the significant differences between the calibrated and uncalibrated figures, as well as the advantages deriving from use of the model. Finally, in the conclusion, we also suggest some other possible areas of application.

2. PRELIMINARY FEATURES: DATA AND INDICATORS

2.1 The Data: The Dutch Time-Budget Surveys

This study uses the Time-Budget Survey (TBO), which was conducted every five years by The Netherlands Institute for Social Research (SCP) and partners; it was repeated with a virtually unchanged format from 1975 to 2005. For the purposes of this article, we only use surveys from four years: 1975, 1980, 2000, and 2005.[2]

In the survey, respondents kept a diary over a seven-day period in October,[3] recording what they were doing (every 15 minutes) and where they were.[4] In addition, respondents were given a detailed core questionnaire in the weeks before they compiled the diary and a shorter final questionnaire in subsequent weeks. Each wave is representative of the entire Dutch population over age 12, but for the purposes of this study, we single out respondents age 18 to 74.

Response rates are generally low in the Netherlands, particularly for time diaries, which are time intensive. They varied between 76 percent in 1975, 54 percent in 1980, and 20 percent (the low point) in 1995 (response formula AAPOR: RR1, see also Mandemakers and Roeters 2014). More recently, various incentives have been offered to increase the return (Dool 2006),[5] with the result that the response rate climbed from 25 percent in 2000 to 37 percent in 2005.[6] To minimize the potential bias, the analysis was weighted to represent the population in terms of gender, age, occupational status, urbanization, size and type of household, and place in the household (Mandemakers and Roeters 2014).

Another difficulty concerns the different level of intensity of an individual's attendance at religious services over the course of a year. Unfortunately, we have no data about monthly variability in attendance in the Netherlands. However, October and November are the months that come closest to the yearly average figure (Harvey 1993), so we believe the result of the calibration is obtained with a reasonable degree of approximation.[7]

As highlighted earlier, to make the comparison, the same activity (here, religious practice) needs to be surveyed using both a diary and a questionnaire. For the diary, we used the code regarding acts of worship, which excludes any such behavior carried out at home. The definition of attendance we adopted for measured presence is "Going to church, attending a humanist gathering etc. *not* at home[8] *in the* survey week."

With regard to the questionnaire, we used two items.[9] The first was a filter question: "Do you regard yourself as belonging to a church community, religion or ideological grouping?" with two answer categories: "yes" and "no." If yes was selected, the second item about frequency of attendance at religious services was applied.[10] This second item—"Do you sometimes go to a church, mosque or other house of prayer? If so, how often do you go on average?"—offered nine answer options: (a) never; (b) less than once a year; (c) once every 7 to 12 months; (d) once every four to six months; (e) once every two to three months; (f) once a month; (g) once every two to three weeks; (h) once a week; and (i) twice a week or more.

2.2 The Indicators: Presence and Frequency

The data typically obtained from diaries and questionnaires are fundamentally different in nature. We first present a simplified description of an indicator derived from diary usage with exclusive reference to religious practice.

After determining the total number of subjects that have to fill in a diary (*N*), we construct subsamples, each one consisting of *N/W* individuals. *W* is the number of subsamples, which is the total period in weeks over which diaries are kept. This period *W* is generally 52, although there are exceptions to the rule, such as in our case where diaries are kept for two weeks ($W = 2$).

The members of the first subsample are asked to fill in a diary of all activities carried out during the first week of the survey period. Similarly, members of the second subsample fill in a diary about the second week, and so on with the acquisition of data of presence/absence covering *W* weeks of the survey period.

The data are then arranged in a matrix with *N/W* rows and *W* columns. The box in the first row and first column is used to record the presence/absence ($x = 1,0$) of the first respondent in the first week of the survey period, the next box in the same row is used for the presence/absence of the second respondent, and so on. In this way, each column contains the attendance data recorded by *N/W* respondents in the corresponding week. Therefore, the ratio $\sum_{i=1}^{N/w} x_{i,1}/N/W$, regarding the first week (first column in the matrix), can be referred directly to a group of individuals, as it shows the proportion of respondents that attended a service during the week in question.

Yet, each row of the matrix contains data taken from diaries filled in by W different respondents in the corresponding W weeks. As a result, the ratio $\sum_{j=1}^{w} x_{1,j}/W$, regarding the first row of the matrix, only shows the proportion of positive events in the row. This observation is useful to understand that each subsample *N/W* is completely independent from the others and there is no connection between these subgroups. In other words, these data are of little practical use. We will return to this issue.

We can now extend the study to the entire matrix by calculating the ratio $P = \sum_{j=1}^{W}\sum_{i=1}^{N/W} x_{i,j}/N$, thereby showing the proportion between positive and possible events. We use the term *measured presence* to refer to the index P (Rossi and Scappini 2014). This idea is not new. Similar definitions have been formulated, for example, "an implicit estimate of the probability of her or his having attended church in «the last seven days»," (Hout and Greeley 1998), the "participation rate . . . of the diary week" (Gershuny 2003); and the "probability of attending on any given week" (Presser and Chaves 2007).

We now examine the practical use of the indicator defined in this way. With regard to the Netherlands, measured presence is 23.06 percent in the years 1975/1980. This means the population attended 23.06 percent of the possible participation events over the survey period. It also means that, on average, each inhabitant attended a religious service in approximately 12 out of the 52 weeks in the year (0.2306 × 52).

However, P is a poor index of information because it does not directly address the respondents as individuals. We only know their share oscillates between a minimum of 23.06 percent for assiduous attendees (52 out of 52 weeks) and a maximum of 100 percent if everybody attends 12 weeks a year. Generally, one would find a combination of the two situations.

This limitation derives from the previously underlined fact that subjects surveyed in different weeks belong to different subsamples. It is therefore not possible to select the part of the population that carries out a given activity within a specific intensity range, as we only know whether each respondent attended the event in the surveyed week and nothing more. To make this distinction we would require access to diaries covering a considerable period of time. For example, if we wanted to select respondents who attended a religious service at least once a month during the past year, we would need to have a diary kept on an annual basis (Scappini 2010). It would clearly be unworkable to use a survey tool of this nature on a large scale.

The typical solution to this problem is to adopt a questionnaire that can determine how frequently each subject attends religious services on a weekly basis over a given period, usually one year. Ideally, if n is the number of weeks in the given period (generally $n = 52$), the $n + 1$ values f_x can be calculated, each of which shows the number of people attending a service x times, with x ranging from 0 to 52. Each ratio f_x/N, in which N is the size of the sample, provides the weekly attendance rate for each single value or group of values of x. This rate can also assume the form of a cumulative attendance rate (here simply cumulative rate) to indicate the proportion of people who attend services at least x times per year, $Cr(x) = \sum_{x}^{52} f_x/N, \forall_x \geq X$.

However, it is unrealistic to ask respondents for such precise data about their attendance at weekly religious services over the course of a year. In general, as in this case, it is preferable to offer a limited number of answer options that correspond more or less directly to frequency intervals.

As a follow-up to this study, we will need to make comparisons and quantify the extent of bias between surveys based on questionnaires and diaries. Although the less detailed data provided by diaries cannot be converted into the more substantial questionnaire data, the reverse process is possible. Similar to prior work (e.g., Gershuny 2003; Hout and Greeley 1998; Presser and Chaves 2007; Presser and Stinson 1998; Woodberry 1998), to make this conversion we add together the products of the number of people and the relative typical attendance frequency *x*—thereby identifying the positive events—and divide the result by the number of possible events: in formal terms, $P(x) = \sum_{x=0}^{52} f_x \cdot x/(N \cdot 52)$. In this instance, we use the term *converted presence*, indicated by P^c.

To make this conversion, we have to tackle an additional problem: identifying the characteristic frequency for each category *x* of the items. For the purposes of this study, the frequency values used will be those indicated in cases where the option refers to a single value, or the central value where there is a range.[11]

We needed to perform two data operations to draw our comparisons. First, we combined the years used, so 1975 and 1980 are referred to as 1980, and 2000 and 2005 are referred to as 2000. Second, we grouped together options *b*, *c,* and *d* in the questionnaire into a single category.[12] In this way, with the answer options ordered by increasing frequency intervals, we reconstructed a correlated monotonic trend of measured presence, which is needed to calibrate the estimates.

To quantify the extent of the bias, we then make comparisons between the presence values with the assumption that the measured presence values are correct and call this index I^p, obtained from $I^p = P^c/P$. We also make comparisons between some values of the cumulative ratio, calculated using questionnaire data, $Cr(x)$, and the relative values obtained from the cumulative density functions that will be calculated with the calibration, which we call $Cc(x)$.[13] We also assume that the calibrated values are correct and call this index I^c, obtained from $I^c = Cr(x)/Cc(x)$.

3. CALIBRATING ESTIMATES

We now address the main focus of this study: calibrating estimates. First, we use data from 1980 to make some comparisons between questionnaires and diaries. As Table 1 shows, the overall population (see last column) reported that they attend services, P^c, to a significantly greater extent than they actually do, verified by the value of measured presence, P. In our case, the former is 29.5 percent and the latter drops to 23.06 percent, with an overestimate of 28 percent (I^b= 1.28).

<Table 1 about here>

This bias is not constant but changes on the basis of the chosen option, roughly following an inverted-U trend. It is lower for the most steadfast attendees (option 7 in the questionnaire), with 4.8 percent of the population claiming to attend services twice a week or more, $P^c = 100.0$ percent. In reality, however, this statement is optimistic: the level of real presence measured using the diary is 87.2 percent, P, with an overestimate of 15 percent, I^b= 1.15. Furthermore, the bias is higher among those who attend services less frequently but regularly, options 4, 5, and 6 in the questionnaire. In these cases, the overestimate ranges from a minimum of 29 to a maximum of 94 percent. Finally, the figure drops again among respondents who attend services more occasionally (option 2) or never (option 1), with notable underestimates of 38 percent (I^b= .62) and 100 percent (I^b= .00), respectively.

Although there is broad scope and variability in terms of errors, respondents are expected to give answers with a minimum level of accuracy reported with a certain degree of honesty (Gershuny 2012:256). In other words, it is reasonable to assume that respondents do not provide values randomly but answer with a certain degree of rationality. Therefore, although the given value on the frequency scale is subject to bias, it is in some way an indicator of "something real." We could also see the strong correlation between measured presence and converted presence as confirmation of this assumption.[14]

Furthermore, measured presence is a continuous indicator, as in theory it could take any of the infinite number of values included in the interval $P[0{,}100]$. As we have seen, this figure cannot be used to make statements about intensity distribution in the population. However, in this case we can obtain more precise information as we can connect measured presence with the values of attendance frequency x stated in the questionnaire. If we establish $X\,\{1{,}2, \ldots{,}7\}$ as a discrete variable with $P_x\,\{p_1{,}p_2, \ldots{,}p_7\}$ and $D_x\{d_1{,}d_2, \ldots{,}d_7\}$, in which 1,2,..,7 are the stated options in the questionnaire, p_i is the measured presence in the specific subpopulation, and d_i is the relative fraction of the population, then $P(X = p_i) \Longrightarrow D = d_i$. We can now express the probability mass function (PMF) for $P(x)$:

$$P(x) = \begin{cases} d_1 & if\ x = p_1 \\ d_2 & if\ x = p_2 \\ & \ldots \\ d_7 & if\ x = p_7 \\ 0 & else \end{cases}$$

This solution has some advantages—it makes it possible to associate the exact presence value for each subgroup, P_x, with the relative fraction of individuals, D_x—but it nevertheless raises three problems. First, as we have seen, the measured presence for each answer option does not generally coincide with the usable figure indicated by converted presence. Second, the levels of P_x may be unstable: as we will see, such values can vary significantly over time. Third, it is extremely unlikely that such broken-down presence data would build up for specific values, as it would actually be distributed along a continuum in which the identified figure is at most an average value.

3.1 Uniform Calibration

The next step is to find a way to interpolate the values of measured presence to be able to solve the aforementioned problems. We will start with an example by considering option (5) in the questionnaire, which corresponds to the code once every two to three weeks and includes respondents who attend services 26 to 17 times a year. As we have seen, this range is at best indicative, as the real frequency figure is much lower. The measured presence of respondents that chose this option is $P(X = 5) = 32.1$ percent (referred to as p_5 here), which corresponds to an attendance rate of 16.7 times a year, a figure lower than the lowest extreme specified in the option.

We will now analyze this situation in depth. Among questionnaire respondents, 5.4 percent chose this option. However, we must consider that the value p_5 is at best an average value, as it is extremely improbable that all individuals who chose this answer behave in an identical manner. There are likely notable variations in their attendance levels: some individuals will attend more occasionally and others more frequently.

Figure 1 shows a simulation of a more realistic hypothetical situation. It might be possible, for example, to distinguish three subgroups among respondents who selected option (5), all made up of the same proportion of individuals (d = 1.8 percent). We can identify a first group in which the average attendance is once a month ($p_{5.1} = 23.1$ percent), a second group with an average of 17 services a year ($p_{5.2} = 32.1$ percent), and a third with an average attendance rate of almost twice a month ($p_{5.3} = 41.0$ percent). This simulation requires two assumptions: first, that no single group consists of 5.4 percent of the sample, but three equal subgroups each consist of 1.8 percent, and second, that the groups are equally spaced ($p_{5.3} - p_{5.2} = p_{5.2} - p_{5.1}$).

<Figure 1 about here>

One could claim that this simulation is too approximate, as it is extremely improbable that individuals who chose this option would actually be distributed into three subgroups. Instead, it is entirely plausible that a continuum of subject distribution can be identified and represented by a function. To establish this function, we must first set its minimum and maximum values. If we only consider answer option (5), these values are unknown: they might range from a minimum of 23.1 percent to a maximum of 41.0 percent, as exemplified, but they may also be different or unsymmetrical. We know nothing about this point.

We now move to option (4), once a month. The same reasoning applied to (5) is also plausible for this option, by first approximating three homogenous subgroups and then using a continuum that can be described by a function. However, the problem of the values used to define the function remains unresolved. Similar reasoning can also be applied to option (6).

Although we cannot identify the values that define the function by working from the single answer options (4), (5), and (6), we can assume it can be calculated in their interlinking continuum. For example, we can imagine respondents selected option (5) when they perceived it as the value that best represents their behavior. In other words, if individuals think they attend services between 17 and 26 times a year, they will choose this option even if the real figure is lower. Similarly, if they think their attendance is around 12 times a year, they will carry out the same reasoning and select option (4). They will probably choose this option even if their attendance is slightly more or less regular, for example, 13 to 16 times a year in the latter case. We do not know the limit value for the transition from (4) to (5), but measured values are somewhat different from stated values, so we know the suggested range is partially hypothetical.

The problem we now have is to find a rule or criterion that allows us to identify a continuum between the calibrated values corresponding to options (4) and (5). The starting point is the aforementioned assumption that respondents choose questionnaire options in a non-arbitrary manner, that is, they refer to their "real behavior" when answering, albeit with approximation. Based on this consideration, we suggest a simple solution.

Our simple solution is to establish the left-hand limit of the distribution of option (5) with the arithmetic mean of the two measured presence values for (4) and (5), $(p_4 + p_5)/2$. *Similarly, we identify the right-hand limit with the arithmetic mean of the two measured presence values for (5) and (6).*[15]

It must be stressed that our underlying assumption of a continuum of subject distribution that can be represented by a function is only valid if P_x is strictly monotonic and therefore only if $p_{x-1} < p_x$, otherwise the behavior we assume to be reasonable would not make much sense.

To further complicate the situation, the real function identified by these two points is unknown: let us assume for now that there is an extremely simple PDF with uniform distribution. We will see how to weaken this assumption. The underlying areas of each identified interval of x are known, as they are simply the fractions of the subjects associated with that option. It follows that $x_i = \left[\frac{p_{i-1} + p_i}{2}, \frac{p_i + p_{i+1}}{2}\right) \Longrightarrow D = d_i$, in which d is now the density of the group that attend services with a measured presence of x_i.

The interpolation of the uniform distribution of x_5 is shown in Figure 1. The left-hand limit, 22.0 percent, is the mean value between p_4 and p_5, calculated from $\frac{11.9 + 32.1}{2}$. The right-hand limit is the mean value between p_5 and p_6, calculated from $\frac{32.1 + 70.3}{2}$ $= 51.2$. Finally, $d_5 = \frac{5.4}{51.2 - 22.0} = 0.2$.

Building on this idea, we now develop a more general formulation:

$$P(x_i) = \begin{cases} \frac{d_1}{(p_1 + p_2)/2 - p_{\min}} & if\ p_{\min} \leq x_1 < \frac{p_1 + p_2}{2} \\ \frac{d_i}{(p_{i+1} - p_{i-1})/2} & if\ \frac{p_{i-1} + p_i}{2} \leq x_i < \frac{p_i + p_{i+1}}{2} \\ \frac{d_7}{p_{\max} - (p_6 + p_7)/2} & if\ \frac{p_6 + p_7}{2} \leq x_7 \leq p_{\max} \end{cases}$$

defined $\forall i \in \{1,2, \ldots,7\}$, in which $P_{min} = [0, p_1]$ and $P_{max} = [p_7, 100]$.

The minimum and maximum values are hypothetical, and in the case of P_{min} can vary between an improbable maximum that coincides with p_1 and a possible minimum of zero. For our study, we chose the intermediate solution, which we feel is most likely, $p_{min} = (0 + p_1)/2$. We followed a similar line of reasoning for $p_{max} = (p_7 + 100)/2$. Although these are conjectured values, they are needed to calculate p_1, p_7, and thereby plot the overall pattern of the function.

With regard to P_{min}, the problem stems from the non-negligible level of measured presence among respondents who stated they were unaffiliated with any denomination. Because these cases have no entry about the frequency of attending services, we cannot associate the value of measured presence with any value in the scale of frequency measured with the questionnaire.[16] Albeit to a much lesser extent, the problem also arises in cases where affiliation with a given denomination was stated, namely for respondents who answered (a) never. Given the demonstrative purposes of this study, we decided to suggest a general solution to the problem. However, one could consider this type of practice as more of a social than a religious activity. By following this last assumption, we could establish $p_1 = 0$.

Otherwise, there is no similar solution for p_{max}, also considering that $p_7 = 87.2$ percent, a value significantly lower than 100 percent. Furthermore, it would be extremely difficult to attend services in all 52 weeks a year, given that many factors (e.g., illness, temporary absences from home) stop even the most assiduous worshippers from achieving this goal. Therefore, in this case the suggested solution seems to be the most reasonable option.

Figure 2 shows the overall result of the reasoning carried out thus far (see uniform PDF). We can now integrate the PDF and establish the relative CDF.[17]

$$F(x_i) = \begin{cases} 0 & if\ x_i < p_0 \\ d_1 \frac{x_i - p_0}{(p_1 + p_2)/2 - p_0} & if\ p_0 \leq x_1 < \frac{p_1 + p_2}{2} \\ \sum_{j=1}^{j=i-1} d_j + d_i \frac{x_i - (p_{i-1} + p_i)/2}{(p_{i+1} - p_{i-1})/2} & if\ \frac{p_{i-1} + p_i}{2} \leq x_i < \frac{p_i + p_{i+1}}{2} \\ \sum_{j=1}^{j=6} d_j + d_7 \frac{x_i - (p_6 + p_7)/2}{p_8 - (p_6 + p_7)/2} & if\ \frac{p_6 + p_7}{2} \leq x_8 < p_8 \\ 1 & if\ x_i \geq p_8 \end{cases} \quad (1)$$

$\forall i \in \{1,2, \ldots,8\}$ in which i is defined from $\frac{p_{i-1} + p_i}{2} \leq x_i < \frac{p_i + p_{i+1}}{2}$, with $p_0 = \frac{0 + p_1}{2}$ and $p_8 = \frac{p_7 + 100}{2}$.

<Figure 2 about here>

The next step is to estimate the proportion of the population that carries out religious practice for specific intensity values, such as the fraction that attends a service at least once a month,[18] and then calculate $Cc(x) = P(X \geq 23.1)$.

To carry out this procedure, we first need to identify the value of i and then calculate the value of $F(x)$ with (1). We then have $i = 5$, as $\frac{p_4 + p_5}{2} \leq x_5 < \frac{p_5 + p_6}{2} = \frac{11.9 + 32.1}{2} \leq x_5 < \frac{32.1 + 70.3}{2} = 22.0 \leq x_5 < 51.2$. It follows that $Cc(x_5) = 100 - \left(69.7 + 5.4 \cdot \frac{2 \cdot 23.1 - 11.9 - 33.1}{70.3 - 11.9}\right) = 30.1$ percent,[19] against 36.6 percent of the estimate from data derived from questionnaires,[20] which corresponds to a value of $I^c = 1.22$.

In this way, we obtained an initial result that is much less subject to bias than that derived solely from using a questionnaire. However, the assumption of uniform distribution is extremely improbable in practice; it is unlikely the PDF pattern would feature breaks at the transition between the different values of $P(x_i)$. A more reasonable solution is that the evolution from $P(x_i)$ to $P(x_{i+1})$ needs to be more progressive. We address this issue in the next section.

3.2 An Improved Model, Linear Calibration

The procedure followed thus far has not been purely illustrative, as it has allowed us to establish the coordinates (x_i, y_i) of some important points $(p_1, p_2, ..., p_7)$. Therefore, given the *abscissa* (p_i), the *ordinate* will be $y_i = \frac{d_i}{(p_{i+1} - p_{i-1})/2}$, $\forall i \in \{2,4, ...,6\}$. For the two tails we have $y_i = \frac{d_i}{(p_i - p_{i-1}) + (p_{i+1} - p_i)/2}$ if $i = 1$ and $y_i = \frac{d_i}{(p_i - p_{i-1})/2 + (p_{i+1} - p_i)}$ if $i = 7$, with $p_0 = \frac{0 + p_1}{2}$ and $p_8 = \frac{p_7 + 100}{2}$. The next step is to produce a linear interpolation starting from the recognized information.

If we now establish $m_i = \frac{y_i - y_{i-1}}{p_i - p_{i-1}}$ and $q_i = \frac{p_i \cdot y_{i-1} - p_{i-1} \cdot y_i}{p_i - p_{i-1}}$, then

$$P(x_i) = \{ m_i \cdot x_i + q_i \quad if \quad p_{i-1} \leq x_i < p_i \qquad (2)$$

with $p_{i-1} < p_i$ by assumption and defined $\forall i \in \{2,3, ...,7\}$.

Note that if x_i is lower than p_1 and higher than p_7, we again apply the uniform model, as these values can only be conjectured and are therefore only useful for the completion of the function. In particular, the ordinates of p_0 and p_8 will be equal to y_1 and y_7, respectively. Figure 3 shows the procedure followed for x_5 (values 32.1 and 0.2).

<Figure 3 about here>

We can now integrate the PDF and establish the relative CDF.[21]

$$F(x_i) = \begin{cases} 0 & if\ x < p_0 \\ d_1 \frac{x_i - p_0}{(p_1 + p_2)/2 - p_0} & if\ p_0 \leq x_1 < p_1 \\ d_1 \frac{(p_1 - p_0)}{(p_1 + p_2)/2 - p_0} + m_2\left(\frac{x_2^2 - p_1^2}{2}\right) + q_2(x_2 - p_1) & if\ p_1 \leq x_2 < p_2 \\ \sum_{j=1}^{j=i-2} d_j + d_{i-1} \frac{(p_{i-1} - p_{i-2})}{p_i - p_{i-2}} + m_i\left(\frac{x_i^2 - p_{i-1}^2}{2}\right) + q_i(x_i - p_{i-1}) & if\ p_{i-1} \leq x_i < p_i \\ \sum_{j=1}^{j=6} d_j + d_7 \frac{x_i - (p_6 + p_7)/2}{p_8 - (p_6 + p_7)/2} & if\ p_7 \leq x_8 < p_8 \\ 1 & if\ x \geq p_8 \end{cases} \quad (3)$$

defined $\forall i \in \{1,2,\ldots,8\}$.

The limits would be defined between p_2 and p_6, as p_1 and p_7 can be calculated with the previously discussed hypothetical assumptions. However, we believe it is also possible to obtain a good approximation for the values between $[p_1, p_2]$ and $[p_6, p_7]$, especially if they are close to p_2 and p_6, respectively. Finally, to obtain a measure of potential error it is sufficient to vary the values of p_0 and p_8. We have now obtained a less approximate description of the PDF pattern, from which we can establish the corresponding CDF (see Figure 4).

[Figure 4 about here]

The proportion of the population that attends a service at least once a month will now be estimated. In a similar way to the previous procedure, we first need to identify the value of i and then calculate the value $F(x)$. We have $I = 5$, as $p_4 \leq x_5 < p_5 = 11.9 \leq x_5 < 32.1$, so for (2), it follows that

$$m_5 = \frac{\frac{5.4}{(70.3-11.9)/2} - \frac{6.3}{(32.1-8.4)/2}}{32.1-11.9} = -.017 \quad \text{and} \quad q_5 = \frac{\frac{32.1\cdot 6.3}{(32.1-8.4)/2} - \frac{11.9\cdot 5.4}{(70.3-11.9)/2}}{32.1-11.9} = .738$$

and for (3)

$$Cc(x_5) = 100 - \left(63.4 + \frac{6.3\cdot(11.9-8.4)}{32.1-8.4} - .017\cdot\left(\frac{23.1^2-11.9^2}{2}\right) + .738\cdot(23.1-11.9)\right) = 30.8$$

percent, which corresponds to a value of $I^c = 1.19$. In this case, there is more limited relative improvement compared to the uniform model.

4. SUMMARY

We will now provide a more complete overview of the situation and recapitulate the advantages of the suggested calibration technique. Table 2 summarizes the main results for 1980 and 2000. For simplicity, we only present two measures of attendance at services, once a month and once every two to three weeks; the other two are the minimum and maximum values that can be calculated with the calibration, also taking into account the figures recorded for 2000.

<Table 2 about here>

We will start with the two extreme values for 1980. In Table 2, we see that only a small fraction of the population, 3 percent, attends services with a relatively high intensity—more than 44 services a year, or an attendance rate of 84.6 percent or more. Assiduous worshippers were already extremely rare in 1980. However, when considering the uncalibrated values, about 25 percent of the population attended a service once a week or more frequently. In this case, the level of overestimate is enormous, with $I^c > 8$.[22] By contrast, a relatively low fraction of the population accounts for occasional or very limited intensity of attendance (about 40 percent attended less than twice a year) compared to the uncalibrated values, where 53 percent of the population stated they very rarely or never attend services.[23] In this case, the level of overestimate is still high, but it is significantly lower than the previous instance (I^c = 1.33).

We now turn to the other two cases included in Table 2. In these situations, we have little to add to what has already been highlighted. We simply reiterate that the cumulative values calculated using questionnaires overestimate the fraction of the population that attends services much more significantly than the calibrated values. Indeed, the I^c values are high among respondents who said they attend a service once or twice a month, 1.19 and 1.16, respectively.

We will now make a diachronic comparison between 1980 and 2000. Note that this comparison is made possible through the calibration of estimates, as data collected with questionnaires would not be directly usable. If we look, for example, at respondents who stated they attend a service at least once a month, there is a measured presence of 11.9 percent in 1980 (see Table 1), which rises to 16.2 percent in 2000 (see Appendix Table A2). The actual behavior of respondents who chose this option changed quite notably at different times, even though the scale of frequency was the same. Conversely, in the diachronic comparison, the measured presence of respondents who said they attend a service once every two to three weeks or every week was similar. In 2000, we see a significant relative drop in the fraction of the population attending religious services at least once a month. In all cases, the change is only partially highlighted by the collected frequency values compared to the calibrated values.

If we consider the group that attends religious services most consistently, the calibrated values feature a 33 percent drop in relative terms (Cr falling from 3.0 to 2.0

percentage points), and the uncalibrated figures show a relative 56 percent drop (Cc plummeting from 24.8 to 11.0 percentage points), with a consistent reduction in the level of overestimate (I^c dropping from 8 to 5). For the population that attends at least once a month, the calibrated values reveal a 47 percent drop in relative terms, falling from 30.8 to 16.2 percentage points, and the uncalibrated figures show a 43 percent drop from 36.6 to 21.0 percentage points, with an increase in the level of overestimate (I^c rising from 1.19 to 1.30). Finally, unlike in 1980, in 2000 a similar fraction of respondents attended fewer than two services a year, around 65 percent of the population in the calibrated values compared to 70.6 percent in the uncalibrated figures, with a drop in the level of overestimate (I^c falling from 1.33 to 1.09).

The raw numbers might not be sufficiently indicative of the extent of bias in the uncalibrated figures, and although we cannot examine the development of secularization in the Netherlands in depth here, we would like to focus on two important aspects of a dimensional and comparative nature. With regard to the first aspect, it should be clear that calibrated figures can be used to measure the extent of religious practice more accurately. In 1980 and in 2000, we see that the real level of attendance at services is much lower than the results typically recorded by questionnaires.

If we analyze what happened in 1980 in more detail, we see that the number of respondents who said they attend services on a regular basis was far greater than the real figure. Paradoxically, the same was true for respondents who said they attend services rarely or never. In other words, even though there was a greater degree of secularization in society than the questionnaires suggested, religion still constituted a socially significant phenomenon, as relatively few people never attended services.

When comparing this situation to the picture in 2000, we see that the drop in religious practice was greater than the questionnaires indicate, on average. However, certain distinctions need to be made. The average overall level of practice fell significantly, especially for respondents attending once a month or more, but there was a smaller decrease in the most regular attendees, that is, those who said they go every week. Conversely, we see a significant increase in relative terms, far greater than the questionnaires suggest, in people who never or hardly ever attend services (see also the drop in I^c from 1980 to 2000).

In short, calibrated and uncalibrated figures offer significantly different pictures. The questionnaires depict a society with a low degree of secularization and a relatively slow but constant and homogeneous shift away from religion. In contrast, the calibrated figures reveal a society in which the process of secularization is further advanced and developing more rapidly in a seemingly less homogeneous and more polarized way.[24]

5. CONCLUSIONS

Questionnaires provide extremely detailed but unreliable estimates, and diaries supply data that are highly precise but extremely poor. Both instruments provide information that is useful and problematic at the same time. We have seen that calibration can be used to overcome these limits, obviating the scarce precision of questionnaires and the scarce detail of diaries. We have also seen that calibration is simple to apply and requires just a few assumptions limited to the monotonicity of the distribution of presence values across a range of attendance frequency response categories and the linearity of the interpolating function. Therefore, the technique is not only recommended for studying attendance at religious services, but it can be applied to any survey in which a given activity was recorded using both a diary and a questionnaire featuring a coherent item that measures the frequency of repetition over a broad time base such as several weeks, a month, or a year.

There are indeed highly diversified fields of application for these two instruments. We will now examine some, but by no means all, of the possible areas of implementation. First, future research could simply extend the application suggested in this study. Indeed, time use surveys often include a questionnaire with frequency questions, alongside diaries; this is seen in studies of free time in England (Kan and Gershuny 2009), Mass attendance in Canada (Brenner 2011), and work in Germany (Otterbach Sousa-Poza 2010). The technique could also be used effectively in studies measuring use of transport. In this case, the need for diaries covering many weeks could be simplified by joint use of diaries and questionnaires (Axhausen et al. 2002; Hanson and Hanson 1981).

Finally, this method could be extended to the medical/psychological sphere, such as studies on the consumption of alcohol (Townshend and Duka 2002) or food (Brunner et al. 2001; Vereecken and Maes 2003). The two instruments are often used interchangeably in these areas. Using them together could increase accuracy and simplify data collection in cases where the analysis needs to be extended over the long term.

This is not the only advantage. With calibration we can make evaluations and comparisons even if the measured presence $P(X = p_i)$ for a given questionnaire option has different values from the useful figures. In this way, it is possible to analyze values that were not envisaged at the planning stage or, within certain limits, special values such as extreme ones, and make comparisons between sources with different characteristics.

In conclusion, this study has proposed—hopefully in a convincing manner—a useful technique for overcoming the limitations of a single data collection instrument that is also easy to apply. In short, it offers a return in terms of data usability in view of the higher cost incurred for obtaining two different measures of a given activity in the same survey.

REFERENCES

Al Baghal, Tarek, Robert F. Belli, A. Lynn Phillips, and Nicholas Ruther. 2014. "What Are You Doing Now? Activity-Level Responses and Recall Failures in the American Time Use Survey." *Journal of Survey Statistics and Methodology* 2(4):519–37.

Axhausen, Kay W., Andrea Zimmermann, Stefan Schönfelder, Guido Rindsfüser, and Thomas Haupt. 2002. "Observing the Rhythms of Daily Life: A Six-Week Travel Diary." *Transportation* 29(2):95–124.

Barta, William D., Howard Tennen, and Mark D. Litt. 2012. "Measurement Reactivity in Diary Research." Pp. 108–123 in *Handbook of Research Methods for Studying Daily Life*, Vol. 23, edited by M. R. Mehl and T. S. Conner. New York: The Guilford Press.

Belli, Robert F. 1998. "The Structure of Autobiographical Memory and the Event History Calendar: Potential Improvements in the Quality of Retrospective Reports in Surveys." *Memory* 6(4):383–406.

Belli, Robert F., and Mario Callegaro. 2009. "The Emergence of Calendar Interviewing: A Theoretical and Empirical Rationale." Pp. 31–52 in *Calendar and Time Diary Methods in Life Course Research*, edited by R. F. Belli, F. P. Stafford, and A. F. Duane. Thousand Oaks, CA: Sage Publications.

Bolger, Niall, Angelina Davis, and Eshkol Rafaeli. 2003. "Diary Methods: Capturing Life as It Is Lived." *Annual Review of Psychology* 54(1):579–616.

Borra, Cristina, Almudena Sevilla, and Jonathan Gershuny. 2013. "Calibrating Time-Use Estimates for the British Household Panel Survey." *Social Indicators Research* 114(3):1211–24.

Brenner, Philip S. 2011. “Identity Importance and the Overreporting of Religious Service Attendance: Multiple Imputation of Religious Attendance Using the American Time Use Study and the General Social Survey.” *Journal for the Scientific Study of Religion* 50(1):103–15.

Brenner, Philip S., and John DeLamater. 2016. “Measurement Directiveness as a Cause of Response Bias Evidence from Two Survey Experiments.” *Sociological Methods & Research* 45(2):348–71.

Bruce, Steve. 2016. “The Sociology of Late Secularization: Social Divisions and Religiosity.” *The British Journal of Sociology* 67(4):613–31.

Brunner, Eric, Daryth Stallone, Maneesh Juneja, Sheila Bingham, and Michael Marmot. 2001. “Dietary Assessment in Whitehall II: Comparison of 7 d Diet Diary and Food-Frequency Questionnaire and Validity against Biomarkers.” *British Journal of Nutrition* 86(3):405–14.

Dool, R. van den 2006. “Veldwerkverslag Tijdbestedingsonderzoek 2005 (Oude stijl). [Field work report Time Use Research 2005 (Old style)].” Hilversum: Intomart GfK bv (http://www.scp.nl/Onderzoek/Bronnen/Beknopte_onderzoeksbeschrijvingen/Tijdsbestedingsonderzoek_TBO).

Eagle, David E. 2011. “Changing Patterns of Attendance at Religious Services in Canada, 1986–2008.” *Journal for the Scientific Study of Religion* 50 (1):187–200.

Frazis, Harley, and Jay Stewart. 2012. “How to Think about Time-Use Data: What Inferences Can We Make about Long- and Short-Run Time Use from Time Diaries?” *Annals of Economic and Statistics* 105/106:231–45.

Gershuny, Jonathan. 2003. *Changing Times: Work and Leisure in Postindustrial*

Society. New York: Oxford University Press.

Gershuny, Jonathan. 2012. "Too Many Zeros: A Method for Estimating Long-Term Time-Use from Short Diaries." *Annals of Economic and Statistics* 105/106:247–70.

Hadaway, C. Kirk, and Penny Long Marler. 2005. "How Many Americans Attend Worship Each Week? An Alternative Approach to Measurement." *Journal for the Scientific Study of Religion* 44(3):307–22.

Hadaway, C. Kirk, Penny Long Marler, and Mark Chaves. 1993. "What the Pools Don't Show: A Closer Look at U.S. Church Attendance." *American Sociological Review* 58(6):741–52.

Hadaway, C. Kirk, Penny Long Marler, and Mark Chaves. 1998. "Overreporting Church Attendance in America: Evidence That Demands the Same Verdict." *American Political Science Review* 63(1):122–30.

Hanson, Susan, and Perry Hanson. 1981. "The Travel-Activity Patterns of Urban Residents: Dimensions and Relationships to Sociodemographic Characteristics." *Economic Geography* 57(4):332–47.

Harms, Teresa, Jonathan Gershuny, Aiden Doherty, Emma Thomas, Karen Milton, and Charlie Foster. 2019. "A Validation Study of the Eurostat Harmonised European Time Use Study (HETUS) Diary Using Wearable Technology." *BMC Public Health* 19(Suppl 2), 455 (https://doi.org/10.1186/s12889-019-6761-x).

Harvey, Andrew S. 1993. "Guidelines for Time Use Data Collection." *Social Indicators Research* 30:197–228.

Hay, Linda Rudin, William M. Hay, and Hugh V. Angle. 1977. "The Reactivity of Self Recording: A Case Report of a Drug Abuser." *Behavior Therapy* 8(5):1004–07.

Hout, Michael, and Andrew M. Greeley. 1998. "What Church Officials' Reports Don't

Show: Another Look at Church Attendance Data." *American Sociological Review* 63(1):113–19.

Jacobs, Jerry A. 1998. "Measuring Time at Work: Are Self-Reports Accurate?" *Monthly Labor Review* 121(12):42–53.

Jansen, Giedo, Nan Dirk De Graaf, and Ariana Need. 2012. "Explaining the Breakdown of the Religion–Vote Relationship in The Netherlands, 1971–2006." *West European Politics* 35(4):756–83.

Juster, F. Thomas, Hiromi Ono, and Frank P. Stafford. 2003. "An Assessment of Alternative Measures of Time Use." *Sociological Methodology* 33(1):19–54.

Kan, Man Yee, and Jonathan Gershuny. 2006. *Infusing Time Diary Evidence into Panel Data: An Exercise in Calibrating Time-Use Estimates for the BHPS*. Colchester: University of Essex (http://www.iser.essex.ac.uk/pubs/workpaps/).

Kan, Man Yee, and Jonathan Gershuny. 2009. "Calibrating Stylised Time Estimates Using UK Diary Data." *Social Indicators Research* 93(1):239–43.

Kan, Man Yee, and Stephen Pudney. 2008. "Measurement Error in Stylized and Diary Data on Time Use." *Sociological Methodology* 38(1):101–32.

Knippenberg, Hans. 2015. "Secularization and Transformation of Religion in Post-War Europe." Pp. 2101–27 in *The Changing World Religion Map*: *Sacred Places, Identities, Practices and Politics*, Vol. 4, edited by S. D. Brunn. New York: Springer.

Korotitsch, William J., and Rosemery O. Nelson-Gray. 1999. "An Overview of Self-Monitoring Research in Assessment and Treatment." *Psychological Assessment* 11(4):415–25.

Lechner, Frank J. 1996. "Secularization in the Netherlands?" *Journal for the Scientific*

Study of Religion 35(3):252–64.

Mandemakers, Jornt J., and Anne Roeters. 2014. “Fast or Slow Food? Explaining Trends in Food-Related Time in the Netherlands, 1975–2005.” *Acta Sociologica* 58(2):121–37.

Otterbach, Steffen, and Alfonso Sousa-Poza. 2010. “How Accurate Are German Work-Time Data? A Comparison of Time-Diary Reports and Stylized Estimates.” *Social Indicators Research* 97(3):325–39.

Presser, Stanley, and Mark Chaves. 2007. “Is Religious Service Attendance Declining?” *Journal for the Scientific Study of Religion* 46(3):417–23.

Presser, Stanley, and Linda Stinson. 1998. “Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance.” *American Sociological Review* 63(1):137–45.

Robinson, John P. 1985. “The Validity and Reliability of Diaries versus Alternative Time Use Measures.” Pp. 33–62 in *Time, Goods, and Well-Being*, edited by F. T. Juster and F. P. Stafford. Ann Arbor: Institute for Social Research, University of Michigan.

Rossi, Maurizio, and Ettore Scappini. 2012. “How Should Mass Attendance Be Measured? An Italian Case Study.” *Quality & Quantity* 46(6):1897–916.

Rossi, Maurizio, and Ettore Scappini. 2014. “Church Attendance, Problems of Measurement, and Interpreting Indicators: A Study of Religious Practice in the United States, 1975–2010.” *Journal for the Scientific Study of Religion* 53(2):249–67.

Scappini, Ettore. 2010. “Daily Diaries in Time Use Surveys: A Solution to Overcome Measurement Problems in Single-Activity Events with Long Characteristic

Rhythms." *Quality & Quantity* 44(5):915–39.

Shiffman, Saul, Arthur A. Stone, and Michael R. Hufford. 2008. "Ecological Momentary Assessment." *Annual Review of Clinical Psychology* 4:1–32.

Townshend, J. M., and T. Duka. 2002. "Patterns of Alcohol Drinking in a Population of Young Social Drinkers: A Comparison of Questionnaire and Diary Measures." *Alcohol & Alcoholism* 37(2):187–92.

Vereecken, Carine A., and Lea Maes. 2003. "A Belgian Study on the Reliability and Relative Validity of the Health Behaviour in School-Aged Children Food-Frequency Questionnaire." *Public Health Nutrition* 6(6):581–88.

Voas, David, and Mark Chaves. 2016. "Is the United States a Counterexample to the Secularization Thesis?" *American Journal of Sociology* 121(5):1517–56.

Walpole, Ronald E., Raymond H. Myers, Sharon L. Myers, and Keying Ye. 2012. *Probability and Statistics for Engineers and Scientists*, 9th ed. New York: Prentice Hall.

Walthery, Pierre, and Jonathan Gershuny. 2019. "Improving Stylised Working Time Estimates with Time Diary Data: A Multi Study Assessment for the UK." *Social Indicators Research* 144(3):1303–21.

Woodberry, Robert D. 1998. "When Surveys Lie and People Tell the Truth: How Surveys Oversample Church Attenders." *American Sociological Review* 63(1):119–22.

ENDNOTES

---

1. Data source: https://easy.dans.knaw.nl/ui/home.
2. We used these years for two reasons. The first is to increase the sample analyzed in order to reduce the instability of the estimates. We thus grouped together the years 1975/1980 and 2000/2005. The second is to use year groupings that are far apart to maximize differences in the comparisons. We thus removed the years 1985, 1990, and 1995 from the analysis.
3. In 2005, for example, the diary was administered October 2 to 8 and October 9 to 15.
4. Respondents are asked to fill out the diaries at the end of each day. However, the interviewers do not verify this; there is a check at the end of the week when the diaries are collected.
5. Respondents who completed the two interviews and presented a complete diary received remuneration in the form of a gift voucher or a donation to a good cause. In 2005, 15 percent of respondents chose the latter option.
6. In 2005, 6,303 addresses were contacted for the initial interview, of which 5,950 were eligible. Of those, 2,892 agreed to take part in the three research phases, and 2,222 were actually interviewed. The clean sample contains 2,188 cases without age selection, consisting of respondents who completed the pre-interview, diary, and post-interview phases.

[7] One added problem is that there is no data collection during periods when major festivals are celebrated, such as Easter and Christmas. However, even if these festivals involved a significant part of the population—assuming the presence value on the two days is 50 percent—the underestimate in the presence value for 1980 would be about 1 percent. In any case, it is improbable that these subjects would be evenly distributed among the different frequencies of Mass attendance. We feel it would be reasonable to position them among respondents with a low level of attendance, therefore introducing a further element of bias into the calculation of $\boldsymbol{Cc} \leq 2.0$.

8. Activity code = 650, place code ≠ 1.

9. Variables v176 and v179 in the questionnaire.

10. We will only refer to churchgoing. Although the Netherlands is a multi-faith country, non-Christian worship still only accounts for a minority of people: 4 percent in 1975 and 6.1 percent in 2005 (estimates from the time-budget surveys).

11. The frequency values we identified with $(x/52)$ and applied to the different answer options correspond to the following weights: (a) 0 times out of 52 = 0.00; (b) 0.5 times = 0.010; (c) from 1.7 to 1 times = 0.026; (d) from 3 to 2 times = 0.048; (e) from 6 to 4 times = 0.096; (f) 12 times = 0.231; (g) from 26 to 17 times = 0.413; (h) and (i) 52 out of 52 = 1.000.

12. The groupings are due to low numbers in some subsamples. In cases in which churchgoing is rare, even a few cases can cause significant variations in estimates. See also Appendix Tables A1 and A3.

13. Regarding this point, analyses of religious behavior generally present the results of a frequency survey through cumulative estimates, a method often used to simplify discussion of results that could otherwise become unnecessarily complex and elaborate.

14. For example, in Table 1 $R = .992$.

15. In other words, we tried to contextualize and validate the fact that discrete distribution is being approximated by continuous distribution. This is a similar operation to the one often called *continuity correction* (Walpole et al. 2012).

16. Regarding this point, the form of the items in a questionnaire tend to reflect the habits of the individuals in the society in which it is administered. In this case, it was assumed that non-affiliates never attend services. In practice, this is almost the case: for 1980 $P = 2.7$ percent, about one service a year. This behavior is similar to respondents who stated some affiliation but chose the option “never” ($P = 2.0$ percent). The situation does not change much for 2000 (see Appendix Tables A1 and A3).

17. We assume that the described functions comply with the properties of a CDF. Although this is obvious, we want to underline that $d_i \geq 0$ and $\sum_1^n d_i = 1$ by definition as the values d_i are normalized. The following manipulation work will simply be the breakdown and reconstruction of the values of d_i.

18. Typically, the rule is to distinguish between regular worshippers (those who go to church at least once a month or more) and irregular or non-worshippers (all others) (see, e.g., Brenner and DeLamater 2016; Eagle 2011; Jansen, De Graaf, and Need 2012; Voas and Chaves 2016). This is not the only solution; there are variations in which regular churchgoers are those who attend a service at least every two weeks (Lechner 1996) or every week (Knippenberg 2015). For discussion about the model, it is not important which solution is adopted. However, we mainly refer to the most widespread one.

19. These and the following calculations were made with a greater number of digits than reported.

20. See Table 1, sum of the values of the specific population for options 4 to 7.

21. For further analysis of note 17, note that $\forall x_i = p_i$, with $i = \{0,1, \ldots,8\}$, linear CDF = uniform CDF.

22. We have given an approximate overestimate, as Cr and Cc refer to different presence figures ($P^c = 100.0$ percent and $P = 84.6$ percent) that are not directly comparable. The observations are made by default, inasmuch as if we could estimate Cc for $P = 100.0$, we would obtain a much lower figure than the 3 percent indicated.

23. "The penumbra of occasional attenders" (Bruce 2016:614) is an aspect that has probably not been explored enough. One possible explanation is the fact that, as we have seen, when a questionnaire is used, it is difficult to estimate with precision the size of the population that attends religious services rarely or never.

[24] Without diminishing the importance of these considerations, note that the inferences we can make about trends in religious practice over time using questionnaires and calibrated values are not so dissimilar in certain respects. However, this convergence is the result of the fact that the trend is extremely pronounced. The advantages that could be derived from the application of the model would probably be even greater if the variations were less evident.

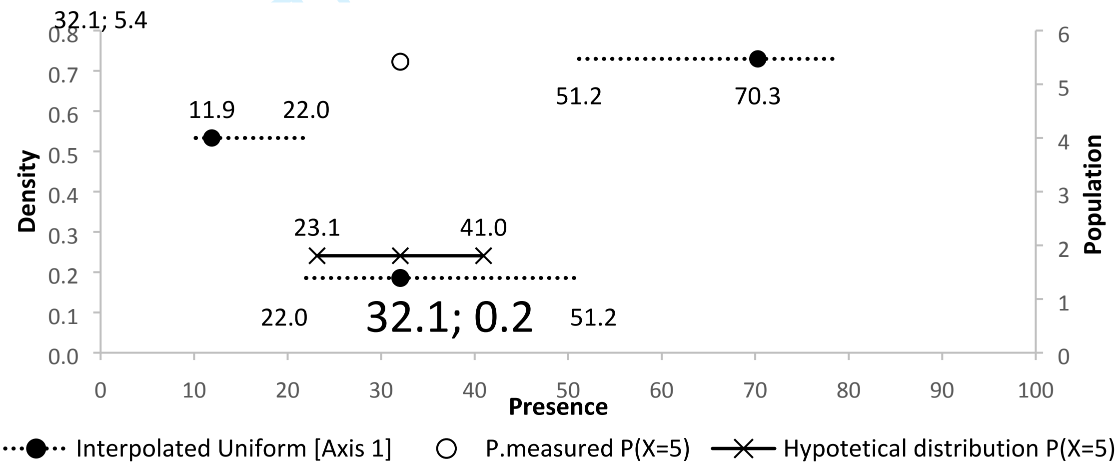

Figure 1. Interpolated Uniform PDF (1980, percent)

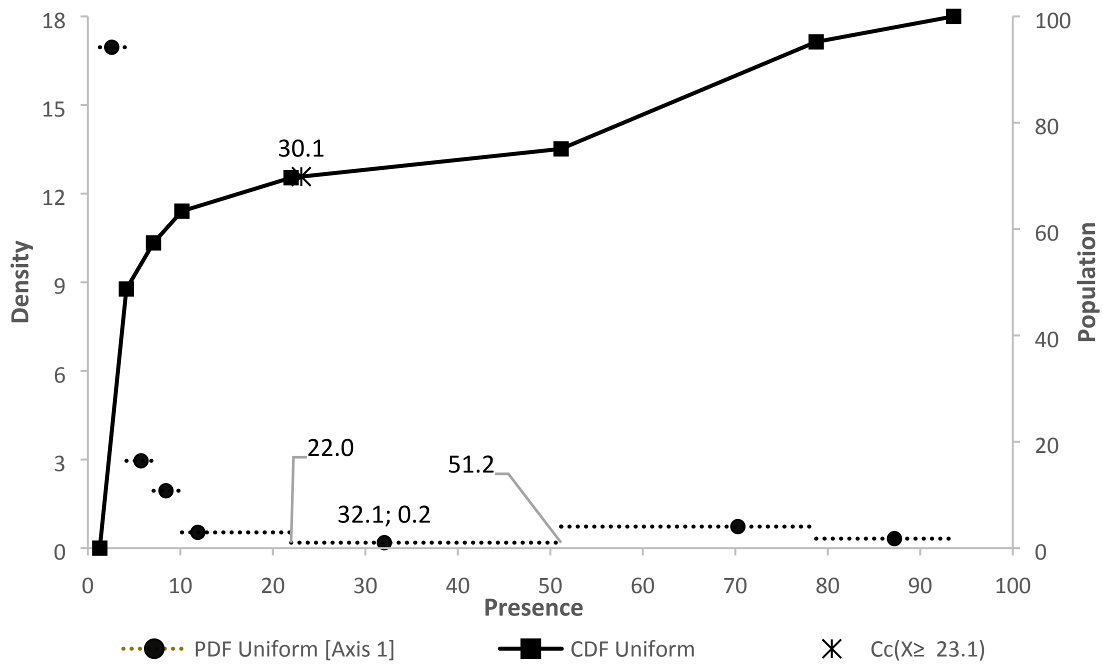

Figure 2. Interpolated Uniform PDF and CDF (1980, percent)

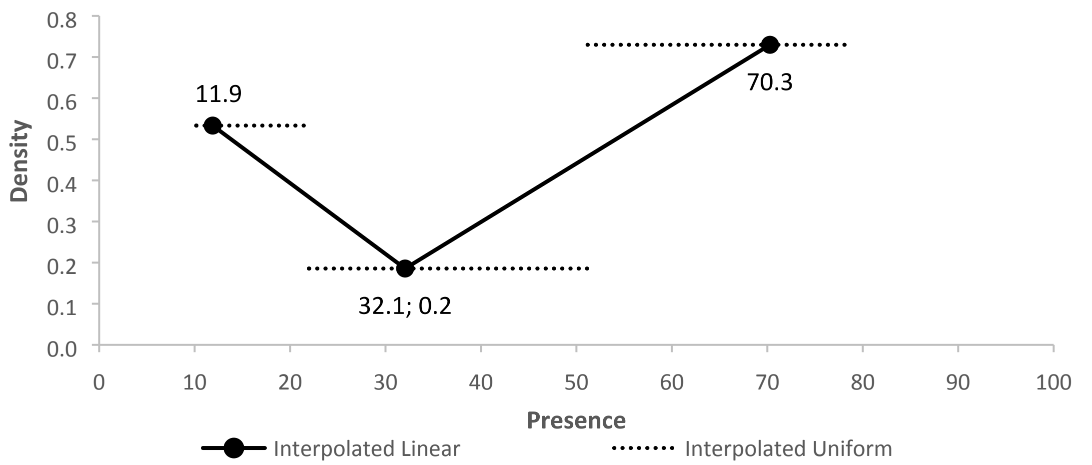

Figure 3. Interpolated Linear PDF (1980, percent)

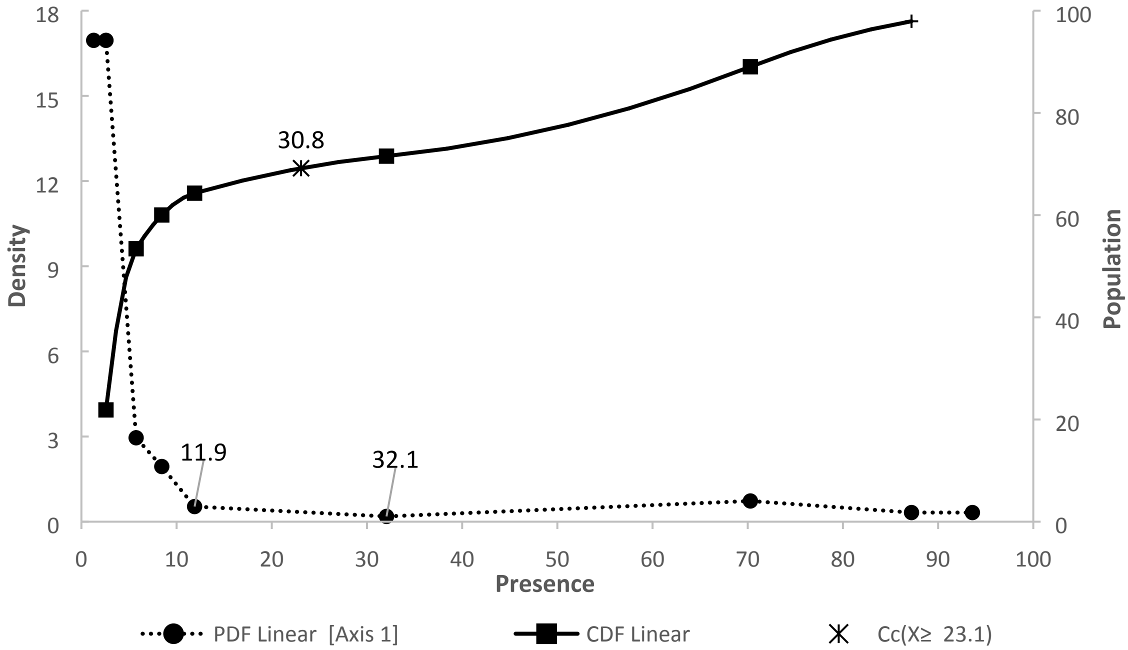

Figure 4. Interpolated Linear PDF and CDF (1980, percent)

Table 1. Presence at religious services (percent) and I^b, by stated frequency of attendance (1980)

Questionnaire options	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*Total*
Converted presence, P^c	0.0	3.5	9.6	23.1	41.3	100.0	100.0	29.5
Measured presence, P	2.6	5.7	8.4	11.9	32.1	70.3	87.2	23.06
Index I^b	0.00	0.62	1.14	1.94	1.29	1.42	1.15	1.28
Specific population, D	48.7	8.7	6.0	6.3	5.4	20.1	4.8	100.0
Cumulative population, Cr	100.0	51.3	42.6	36.6	30.3	24.9	4.8	
(Unweight N)	(1,643)	(293)	(216)	(229)	(195)	(613)	(131)	(3,320)

Note: 1 = never; 2 = less than once a year, once every 7/12 months, once every 4/6 months; 3 = once every 2/3 months; 4 = once a month; 5 = once every 2/3 weeks; 6 = once a week; 7= twice a week or more.

Table 2. Cumulative population (percent) and values of I^c, by frequency of attendance at religious services (1980 and 2000)

Number of annual services (converted and measured presence)	≤ 2.0 (3.8)[a]	≥ 12.0 (23.1)[b]	≥ 21.5 (41.3)[c]	≥ 44.0 (84.6)[d]
Years 1975/1980				
Cumulative population, ***Cr***	53.1	36.6	30.3	24.9

Calibrated cumulative population, $\boldsymbol{Cc}$	39.8	30.8	26.1	3.0
Index $\boldsymbol{I^c}$	1.33	1.19	1.16	>8
Years 2000/2005				
Cumulative population, $\boldsymbol{Cr}$	70.6	21.0	15.2	11.0
Calibrated cumulative population, $\boldsymbol{Cc}$	64.6	16.2	12.4	2.0
Index $\boldsymbol{I^c}$	1.09	1.30	1.23	>5

[a]In this case, the Cr figures were calculated by merging the original answers, not the aggregated answers shown in Table 1. See Appendix Tables A1 and A3, answers (a), (b) and (c).
[b]Option corresponding to the stated frequency in the questionnaire "once a month".
[c]Option corresponding to the stated frequency "once every 2 to 3 weeks".
[d]In this case, Cr was calculated by selecting the answer options corresponding to the stated frequency in the questionnaire (6 = once a week and 7 = twice a week or more), in which $P^c = 100.0$ percent.

APPENDIX

Table A1. Presence at religious services (percent) and I^b, by stated frequency of attendance (1980)

Questionnaire options	Only non-affiliates	Affiliates and answer (a)	*Total*	Affiliates and answer: (b)	(c)	(d)	*Total (b+c+d)*
Converted presence, $\boldsymbol{P^c}$	0.0	0.0	*0.0*	1.0	2.6	4.8	*3.5*
Measured presence, $\boldsymbol{P}$	2.7	2.0	*2.6*	5.1	0.8	9.9	*5.7*
Index $\boldsymbol{I^b}$	0.00	0.00	*0.0*	0.19	3.28	0.48	*0.62*
Specific population, $\boldsymbol{D}$	42.7	6.0	*48.7*	0.8	3.6	4.3	*8.7*
(Unweight *N*)	(1,466)	(177)	*(1,643)*	(23)	(108)	(162)	*(293)*

Note: (a) never; (b) less than once a year; (c) once every 7/12 months; (d) once every 4/6 months.
Option 1 in Table 1 is the aggregation of non-affiliates and affiliates and answer (a).
Option 2 in Table 1 is the aggregation of affiliates and answers (b), (c) and (d).

Table A2. Presence at religious services (percent) and I^b, by stated frequency of attendance (2000)

Questionnaire options	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*Total*
Converted presence, P^c	0.0	2.8	9.6	23.1	41.3	100.0	100.0	14.9
Measured presence, P	0.8	4.5	8.8	16.2	32.0	69.7	86.9	12.0
Index I^b	0.00	0.63	1.09	1.42	1.29	1.44	1.15	1.24
Specific population, D	62.3	12.1	4.6	5.8	4.1	7.6	3.4	100.0
Cumulative population, Cr	100.0	37.7	25.6	21.0	15.2	11.0	3.4	
(Unweight *N*)	(2,149)	(416)	(163)	(184)	(146)	(271)	(109)	(3,438)

Note: 1 = never; 2 = less than once a year, once every 7/12 months, once every 4/6 months; 3 = once every 2/3 months; 4 = once a month; 5 = once every 2/3 weeks; 6 = once a week; 7= twice a week or more.

Table A3. Presence at religious services (percent) and I^b, by stated frequency of attendance (2000)

Questionnaire options	Only non-affiliates	Affiliates and answer (a)	*Total*	Affiliates and answer: (b)	(c)	(d)	*Total (b+c+d)*
Converted presence, $\boldsymbol{P^c}$	0.0	0.0	*0.0*	1.0	2.6	4.8	*2.8*
Measured presence, $\boldsymbol{P}$	0.7	2.0	*0.8*	1.5	3.2	8.9	*4.5*
Index $\boldsymbol{I^b}$	0.00	0.00	*0.00*	0.66	0.82	0.54	*0.63*
Specific population, $\boldsymbol{D}$	56.4	5.9	*62.3*	3.5	4.8	3.8	*12.1*
(Unweight *N*)	(1,940)	(209)	*(2,149)*	(114)	(161)	(141)	*(416)*

Note: (a) never; (b) less than once a year; (c) once every 7/12 months; (d) once every 4/6 months.
Option 1 in Table A2 is the aggregation of non-affiliates and affiliates and answer (a).
Option 2 in Table A2 is the aggregation of affiliates and answers (b), (c) and (d).