Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis

Sindre Rolstad, PhD,* John Adler, PhLic, Anna Rydén, PhD
AstraZeneca R&D, Mölndal, Sweden

ABSTRACT

Background: Response burden is often defined as the effort required by the patient to answer a questionnaire. A factor that has been proposed to affect the response burden is questionnaire length, and this burden is manifested in, for example, response rate. Even though response burden is frequently mentioned as a reason for abridging questionnaires, evidence to support the notion that shorter instruments are preferable is limited. Objectives: This study aimed to accumulate, analyze, and discuss evidence regarding the association between response burden, as measured by response rate, and questionnaire length. Methods: A systematic literature review and meta-analysis of studies reporting response rates in relation to questionnaire length was performed. A Cochran-Mantel-Haenszel test stratified by study using the Breslow-Day test was undertaken to investigate homogeneity of the odds ratios. Results: Thirty-two reports were identified, of which 20 were eligible for inclusion in the meta-analysis. Response rates were lower for longer questionnaires, but because the P value for test of homogeneity was P = 0.03, this association should be interpreted with caution because it is impossible to separate the impact of content from length of the questionnaires. Conclusion: Given the inherently problematic nature of comparing questionnaires of various lengths, it is preferable to base decisions on use of instruments on the content rather than the length per se. Keywords: content validity, meta-analysis, patient-reported outcomes, questionnaire length, response rates.

Introduction

There is no consensus on how to define response burden. Factors such as questionnaire length, density of sampling, cognitive load required completing the survey, and layout and interface of the reporting format have been suggested to affect the strain on the patient [1]. Other similar concepts are respondent burden, response fatigue, and subject burden. Increased response burden has been proposed to result in lower response rates, reduced completion, and reduced data quality [2,3]. Response burden may be particularly problematic in demographic groups such as the severely ill, older individuals, and children [4]. A strong focus has been on questionnaire length, and, consequently, potential response burden is frequently a rationale for reducing the number of items in existing questionnaires (e.g., the short version of the Short-Form Health Survey [SF-36], SF-12 [5]) and is also driving development of questionnaires with a minimum of items. Lengthy questionnaires have been mentioned as a general obstacle in clinical practice [6] and as an argument for limiting the overall number of administrations of an instrument. Some studies have even proposed that a single item is preferable to reduce response burden [7]. Techniques such as item response theory, computer adaptive testing, skip sequencing, and item banking may offer reduced response burden because items may be more tailored to meet patients’ response patterns compared to traditional paper and pencil forms. Even though response burden is frequently mentioned as a reason for reducing the number of items and total administration time, there may not be strong empirical evidence to support the notion that comparatively shorter instruments are preferable. Increased regulatory demands on patient-reported outcomes (PROs) [1] (e.g., acuity of recall accuracy, content validity, and preference for the use of an electronic reporting format) may have resulted in increased sampling density (frequency of measurement). Thus, there is an augmented need to understand whether there is an association between questionnaire length and response burden.

In this review, current evidence concerning response burden in relation to questionnaire length is compiled and analyzed and unresolved issues are exposed, highlighting research priorities for the future.

Methods

Original studies evaluating the relationship between questionnaire length and response rate were identified by systematic searches of PubMed and EMBASE. Studies fulfilling the following criteria were included:

- The relationship between questionnaire length and response rate was investigated.
- The sample size was sufficiently large to provide a relatively precise estimate of the relationship.
- The study design was appropriate for assessing the relationship between questionnaire length and response rate.

A systematic literature review and meta-analysis were performed. A Cochran-Mantel-Haenszel test stratified by study using the Breslow-Day test was undertaken to investigate homogeneity of the odds ratios.
The systematic search for studies on the relationship between questionnaire length and respondent burden yielded 4861 records of potentially relevant studies (Table 1). After screening records and obtaining copies of the reports considered useful for further inspection, a total of 21 reports, 2 meta-analyses and 19 seminal articles were found to match the eligibility criteria. Review of relevant references resulted in the addition of a further 11 seminal articles. Previous relevant meta-analyses on the relationship between questionnaire length and respondent burden are discussed in the following. Seminal articles not included in the previous reviews were subsequently grouped on the basis of whether a significant association between respondent burden, as assessed by response rates, and questionnaire length was reported. Two studies reported a significant association, whereas 12 did not. Significance was defined using authors’ prespecified criteria; in general, a P value <0.05 was used to determine significance. One exception was Jenkinson et al. [10], who required a 10% difference in response rates to be considered significant. Studies investigating the effect of questionnaire length on response rates are also depicted in Table 2. Thereafter, three studies are presented that used methods to obtain information directly from patients regarding respondent burden. Any measure indicating respondent burden, either directly or indirectly, is stated, as is any indication of data quality, if disclosed. The literature review concludes by a meta-analysis of the potential association between response rates and questionnaire length.

In total, 24 of 25 studies reporting response rates compared the effect of questionnaire length across groups; one study was excluded for not reporting response rates [11]. The most commonly reported indication of questionnaire length was pages. Of the 24 eligible studies, 7 did not report the number of items [12–18], whereas 4 did not report the number of pages [19–22]. Accordingly, the number of pages was used as the measure of questionnaire length in the meta-analysis, resulting in an exclusion of the studies only reporting the number of items. All of the 20 studies included in the meta-analysis were cross-sectional, and no more than one measurement was carried out per study participant.

### Statistics

All studies included in the meta-analysis included questionnaires defined by the authors of that report as shorter and longer. These definitions of relative length were used to dichotomize the studies into shorter and longer questionnaires. The difference in response rates between shorter and longer questionnaires was analyzed using a Cochran-Mantel-Haenszel test stratified by study [8]. This is an analysis of the relationship between response rate and length of questionnaires, controlled for study. The Cochran-Mantel-Haenszel estimate of the common odds ratio (OR) over all strata is presented together with the corresponding 95% confidence interval (CI). The Breslow-Day test for homogeneity of the ORs in the studies is also presented [9]. All statistical analyses were performed using SAS software (version 8.2, SAS Institute, Cary, NC) and the R computing environment (version 2.12.0, The R Foundation, http://www.r-project.org).

### Previous meta-analyses

Two previous reviews were found to contain information regarding the potential association between respondent burden and questionnaire length [23,24]. A systematic review of randomized,

<table>
<thead>
<tr>
<th>Search terms</th>
<th>Initial hits</th>
<th>Relevant hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject/patient/response/respondent/testing/number of items + burden</td>
<td>16/98/38/111/8/110</td>
<td>0/1/1/0/0/0</td>
</tr>
<tr>
<td>“Response fatigue”</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>“Burden of report”</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Questionnaire/survey/scale * length + burden</td>
<td>118/645/119</td>
<td>0/0/0</td>
</tr>
<tr>
<td>“Questionnaire length” / “survey length” / “scale length” / “number of items” + response rate</td>
<td>391/451/42/233</td>
<td>2/4/0/0</td>
</tr>
<tr>
<td>“Questionnaire length” / “survey length” / “scale length” / “number of items” + compliance</td>
<td>203/484/146/128</td>
<td>8/0/0/0</td>
</tr>
<tr>
<td>Extensive + questionnaire/survey/scale</td>
<td>187/341/17</td>
<td>0/0/0</td>
</tr>
<tr>
<td>Brief / short / abridge + length + questionnaire/survey/scale</td>
<td>188/149/380</td>
<td>0/0/0</td>
</tr>
<tr>
<td>Long + short / brief / abridge + questionnaire/survey/scale</td>
<td>3/11</td>
<td>0/0/0</td>
</tr>
<tr>
<td>“Concise + questionnaire/survey/scale”</td>
<td>3/21/0</td>
<td>0/0/0</td>
</tr>
<tr>
<td>“Lengthy + questionnaire/survey/scale”</td>
<td>17/0/0</td>
<td>0/0/0</td>
</tr>
<tr>
<td>“Comprehensive + questionnaire/survey/scale” + brief</td>
<td>5/22/5</td>
<td>0/0/0</td>
</tr>
<tr>
<td>“Comprehensive + questionnaire/survey/scale” + short / brief / abridge</td>
<td>22/41/4</td>
<td>0/0/0</td>
</tr>
<tr>
<td>“Long + questionnaire/survey/scale”</td>
<td>14/54/41</td>
<td>4/0/0</td>
</tr>
<tr>
<td>Relevant articles retrieved from relevant hits</td>
<td>n/a</td>
<td>11</td>
</tr>
<tr>
<td>Total no.</td>
<td>4861</td>
<td>32</td>
</tr>
</tbody>
</table>

Relevant hits do not include duplicates of previously listed search terms. n/a, not applicable.

---

**Table 1 – Search terms, number of hits, and number of relevant hits for searches performed in PubMed and EMBASE.**
<table>
<thead>
<tr>
<th>Study</th>
<th>Setting and design</th>
<th>Topic</th>
<th>Questionnaires shorter and longer</th>
<th>Pages shorter</th>
<th>Pages longer</th>
<th>Received/sent (RR %), shorter questionnaire</th>
<th>Received/sent (RR %), longer questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beebe et al., 2010 [35]</td>
<td>Unaffiliated, randomized postal survey</td>
<td>Functional gastrointestinal disorders</td>
<td>Two versions of the Talley Bowel Disease Questionnaire</td>
<td>2</td>
<td>4</td>
<td>169/379 (44.6)</td>
<td>194/401 (48.4)</td>
</tr>
<tr>
<td>Coast et al., 2006 [19]</td>
<td>Affiliated, randomized postal survey</td>
<td>Patient satisfaction with care</td>
<td>Ad hoc without and with more scenarios to consider</td>
<td>16*</td>
<td>36</td>
<td>103/121 (85.1)</td>
<td>99/119 (83.2)</td>
</tr>
<tr>
<td>Cunningham et al., 1999 [20]</td>
<td>Affiliated, randomized postal survey</td>
<td>Alcohol</td>
<td>Brief graduated frequency measure and Timeline Followback</td>
<td>30*</td>
<td>36*</td>
<td>25/49 (51)</td>
<td>10/45 (22)</td>
</tr>
<tr>
<td>de Marco et al., 1999 [31]</td>
<td>Affiliated, randomized postal survey</td>
<td>Asthma</td>
<td>Ad hoc without and with questions of costs</td>
<td>1</td>
<td>2</td>
<td>64/142 (45)</td>
<td>53/140 (38)</td>
</tr>
<tr>
<td>Dirmaier et al., 2007 [34]</td>
<td>Affiliated, randomized postal survey</td>
<td>Psychiatry</td>
<td>Ad hoc without and with standardized psychiatric outcome measures</td>
<td>13</td>
<td>23</td>
<td>1069/1948 (54.9)</td>
<td>960/1877 (51.1)</td>
</tr>
<tr>
<td>Dorman et al., 1997 [26]</td>
<td>Affiliated, randomized postal survey</td>
<td>Quality of life</td>
<td>EuroQol and SF-36</td>
<td>4</td>
<td>8</td>
<td>905/1125 (75)</td>
<td>849/1128 (80)</td>
</tr>
<tr>
<td>Eaker et al., 1998 [27]</td>
<td>Unaffiliated, randomized postal survey</td>
<td>Medical history and risk factors</td>
<td>Ad hoc</td>
<td>11</td>
<td>15</td>
<td>511/1000 (51.1)</td>
<td>464/1000 (46.4)</td>
</tr>
<tr>
<td>Hoffman et al., 1998 [29]</td>
<td>Affiliated, randomized postal survey</td>
<td>Personal and family medical history</td>
<td>Ad hoc</td>
<td>4</td>
<td>16</td>
<td>167/452 (37)</td>
<td>340/1029 (33)</td>
</tr>
<tr>
<td>Jacoby, 1990 [13]</td>
<td>Affiliated and unaffiliated patients, randomized postal survey</td>
<td>Patient satisfaction with care</td>
<td>Ad hoc</td>
<td>8</td>
<td>16</td>
<td>678/1000 (67.8)</td>
<td>684/1000 (68.4)</td>
</tr>
<tr>
<td>Jenkinson et al., 2003 [10]</td>
<td>Affiliated, randomized postal survey</td>
<td>Patient satisfaction with care</td>
<td>Two versions of the Picker Patient Experience</td>
<td>4</td>
<td>12</td>
<td>396/721 (67.7)</td>
<td>362/724 (63.7)</td>
</tr>
<tr>
<td>Kalantar and Talley, 1999 [36]</td>
<td>Unaffiliated, randomized postal survey</td>
<td>Gastrointestinal</td>
<td>Ad hoc</td>
<td>1</td>
<td>7</td>
<td>166/220 (75.6)</td>
<td>150/220 (68.2)</td>
</tr>
<tr>
<td>Kaplan and Cole, 1970 [21]</td>
<td>Unaffiliated, randomized postal survey</td>
<td>Reproductive history</td>
<td>Ad hoc</td>
<td>1*</td>
<td>5*</td>
<td>162/219 (73.9)</td>
<td>167/217 (76.9)</td>
</tr>
<tr>
<td>Kelly et al., 2010 [37]</td>
<td>Unaffiliated, randomized postal survey</td>
<td>Cancer</td>
<td>Ad hoc</td>
<td>10</td>
<td>16</td>
<td>368/573 (64.2)</td>
<td>340/577 (58.9)</td>
</tr>
<tr>
<td>Mond et al., 2004 [15]</td>
<td>Unaffiliated, randomized postal survey</td>
<td>Health and well-being</td>
<td>Ad hoc</td>
<td>8</td>
<td>14</td>
<td>194/401 (48.4)</td>
<td>230/401 (57.4)</td>
</tr>
<tr>
<td>Nagata et al., 1995 [16]</td>
<td>Unaffiliated, randomized postal survey</td>
<td>Medical history and risk factors</td>
<td>Ad hoc</td>
<td>1</td>
<td>2</td>
<td>50/100 (50)</td>
<td>157/300 (52.3)</td>
</tr>
<tr>
<td>Ronckers et al., 2004 [32]</td>
<td>Unaffiliated, randomized postal survey</td>
<td>Nasopharyngeal radium irradiation</td>
<td>Ad hoc</td>
<td>8</td>
<td>12</td>
<td>77/100 (77)</td>
<td>67/100 (67)</td>
</tr>
<tr>
<td>Salisbury et al., 2005 [22]</td>
<td>Affiliated, randomized postal survey</td>
<td>Patient satisfaction with care</td>
<td>The Short Questionnaire for Out-of-Hours Care and McKinley’s Questionnaire for Out-of-Hours Care</td>
<td>7*</td>
<td>36</td>
<td>342/748 (45.7)</td>
<td>234/558 (41.9)</td>
</tr>
</tbody>
</table>

(continued on next page)
controlled trials undertaken to identify approaches that increase response rates of postal surveys concluded that a response was more likely when comparatively shorter questionnaires were used (OR 1.86; 95% CI 1.55–2.24) [23,24]. In this meta-analysis, questionnaires were dichotomized into shorter and longer on the basis of the individual study authors’ definitions. As pointed out by others [25], many of the trials included were outdated and on topics unrelated to health care, for example, marketing surveys. Consequently, these findings have little clinical applicability. When only published reports on a health-related subject cited in the review were included, longer questionnaires were found to result in decreased response rates in 4 of 11 studies [12,26–28]. In the remaining seven studies, there was no evidence of increased burden as measured by response rate [13,14,16–18,21,29]. It should be noted that not all of these studies compared different versions of the same questionnaire, and some compared different questionnaires altogether (e.g., SF-36 vs. EuroQol (EQ-5D) [26]). Ten of these studies were eligible for inclusion in this meta-analysis. This literature review included observational studies and was not restricted to randomized studies.

The potential relationship between response rate and number of items included in a questionnaire was investigated in another review [24]. The number of items was stated in 125 of the studies included and ranged from 1 to 361 (mean 28.1; SD 51). According to authors, the correlation between the number of items and total response rate was weak ($r = 0.29$). The authors concluded that high response rates may be achieved even with longer questionnaires (e.g., 15–20 pages). Evidence of the potential relationship between response rate and questionnaire length not included in the reviews cited previously [23,24] is presented.

### Studies reporting a significant association between burden and length

The relationship between attrition rate and questionnaire length was examined in a postal survey of clinical patients [20]. Participants were randomly assigned to receive either a shorter (brief graduated frequency measure) (n = 49) or a longer (Timeline Followback [TLFB]) (n = 46) questionnaire assessing alcohol consumption. The response rate for the TLFB was 22% (10/45) compared with 51% (25/49) for the relatively shorter questionnaire ($\chi^2 = 6.5, df = 1, P = 0.01$). No differences in demographic characteristics were found for the participants receiving either questionnaire. The authors concluded that shorter questionnaires may be preferable to longer ones in postal surveys.

Chlan [11] compared response burden indirectly by administering two anxiety instruments, the 20-item Spielberger State Anxiety Inventory (SAI) and a visual analog scale for anxiety (VAS-A), to all patients (n = 200) receiving mechanical ventilator support. A VAS consists of one single item where level of agreement to a statement is indicated by making a mark along a continuous line between two end points. Because all patients completed the VAS-A and eight did not complete the SAI, the authors concluded that the VAS-A was less burdensome than the SAI for patients receiving mechanical ventilator support.

### Studies not reporting a significant association between burden and length

A study conducted by Victor [30] investigated the feasibility of administering questionnaires by mail to elderly patients upon hospital discharge. Patients (n = 300) were randomized to receive a questionnaire of either 4 or 12 pages in length on the subject of disability and health status. Response rates among survivors were comparable between the groups; 87% (117/135) completed the shorter questionnaire, whereas 86% (118/138) returned the longer one. Similar data quality was reported for both groups.
The influence of length of a questionnaire on the topic of asthma on response rate was assessed in a pilot study [31]. Two random samples of healthy respondents received either a one-page (n = 150) or a two-page (n = 150) questionnaire. The response rate for the two-page questionnaire was 7% lower than that for the one-page version (38% vs. 45%). The difference was not statistically significant (95% CI –1% to 15%).

An epidemiological study pilot-tested (n = 200) the effect of questionnaire length on response rate before selecting a questionnaire for distribution to a larger study cohort [32]. Although the results were not statistically significant (P = 0.12; 95% CI for the difference in proportion −2% to 22%), the response rate increased from 67% to 77% when an eight-page questionnaire was administered compared to one of 12 pages in length.

Subar et al. [33] administered two food frequency questionnaires by mail to a control group (n = 900) in a cancer screening trial. The authors reported no differences with respect to response rate and data quality when comparing data for those who were randomized to receive either the 16-page Prostate, Lung, Colorectal, and Food Frequency Questionnaire (PLO FFQ) or the 36-page Diet History Questionnaire (DHQ) (81.6% vs. 81.9%). Also, 58% found the DHQ to be very easy to complete, whereas the corresponding figure for the PLO FFQ was 50%. The authors reported that the mean proportion of missing or uninterpretable responses did not differ by questionnaire type.

A study by Jenkinson and colleagues [10] compared two versions of various lengths of the Picker Patient Experience (PPE) questionnaire. The study included patients recently discharged from hospital (n = 1445) who were randomized to receive either a 12- or a 4-page version of the PPE questionnaire. The authors found no difference in response rate; 63.7% returned the longer questionnaire, whereas 67.7% returned the shorter one. Also, the data quality and psychometric properties of the PPE did not differ between the two arms of the trial.

An epidemiological study by Mond et al. [15] on the effect of questionnaire length and response rate concluded that little is gained by decreasing questionnaire length. The overall response rate was significantly higher for participants (n = 401) who had been mailed a 14-page questionnaire (57.4%) compared to those (n = 401) who had received an eight-page version, which excluded some of the domains found in the longer versions (48.4%) (OR 1.44; 95% CI 1.09–1.9; P = 0.01).

Salisbury et al. [22] compared the newly developed seven-item Short Questionnaire for Out-of-Hours Care with a previously validated, longer 36-item questionnaire. Questionnaires were mailed to patients (n = 1906) contacting an out-of-hours general practitioner cooperative. The authors found no significant difference between the overall response rates obtained from the short or long questionnaires (45.7% vs. 41.9%; P = 0.17). More patients in the short questionnaire group, however, completed all items (43.0% vs. 36.4%; P = 0.01).

In another study that sought to identify differences in completion rates in relation to questionnaire length, 36- and 16-item versions of a questionnaire were mailed to affiliated patients on the subject of access to dermatology secondary-care services [19]. The authors reported that 53% of the sample (n = 240) agreed to participate and that the response rate for the longer questionnaire was 83.2% compared to 85.1% for the comparatively shorter questionnaire (difference, 1.9%; 95% CI −7.3 to 11.2; P = 0.68). The authors also reported that more patients completed all items in the shorter questionnaire, but found this to be attributable to differences in items rather than length per se.

Dirmaier et al. [34] investigated whether abridging a longer questionnaire would increase response rates in a postal follow-up survey 1 year after inpatient psychotherapeutic treatment. Patients (n = 3825) received either a prepaid monetary incentive or none, and a 13- or 23-page questionnaire. Although a monetary incentive increased the response rate by 7.3% (95% CI 2.6–11.9), receiving the relatively shorter questionnaire resulted in a nonsignificant difference in response rate of 3.7% (95% CI 0.9–8.3, P = 0.11); the shorter questionnaire was returned by 51.1% of the patients and the longer one by 54.9%. The authors reported no significant associations between questionnaire length and the mean number of missing values when adjusting for age, sex, and diagnosis.

A recent postal survey sought to investigate response burden and data quality indirectly by comparing response rates [25]. A cohort of 40 patients was randomized to receive either a two- or a four-page long version of the Talley Bowel Disease Questionnaire (TBDQ). There were no significant differences (P = 0.29) in response rates between those who had received the long (194/401, 48.4%) and those who had received the short (169/379, 44.6%) version of the TBDQ. Furthermore, the study did not report any differences between the questionnaires with respect to partial responses or time to submit the questionnaire.

Another epidemiological study evaluated the effects of lottery incentive and questionnaire length on response rates when used in isolation or combined [36]. A random sample (n = 440) of individuals was by chance assigned to be given or not given an instant lottery ticket and a seven- or one-page questionnaire. The distribution of the variable conditions, length, and lottery incentive was equal across groups. The final response rates were highest among those completing the shortest questionnaire (75.6% vs. 68.2%; P = 0.08) but not significantly so. The possibility of receiving a completed questionnaire without any follow-up reminders was significantly associated with the lottery incentive (P = 0.03) but not with the length of the questionnaire.

Patients with prostate or colon cancer (n = 1200) were randomly selected from a cancer registry to receive either a 10- or a 16-page questionnaire [37]. The response rate of the patients receiving the comparatively shorter questionnaire was 64.2%, and for the longer questionnaire, it was 58.9% (OR 0.79; P = 0.06). It was concluded that the length of the questionnaire did not affect the response rate significantly, even though there was a trend. No differences between the two questionnaires were found regarding data quality.

Studies including patient input
Hassan and Weymuller [38] tested which instrument was preferred by patients with head and neck cancer (n = 75). Patients were administered the nine-item disease-specific University of Washington-Quality of Life (UW-QOL) head and neck questionnaire and the 136-item generic Sickness Impact Profile (SIP). The authors reported that 97% of patients favored the shorter UW-QOL over the SIP when asked an open-ended question regarding preference. It should be noted, however, that the instruments used for comparison differed markedly in terms of content.

In another study [39] that included patient (n = 114) input, patients with multiple sclerosis (MS) were asked to pick off which of three questionnaires—EQ-5D (five items), SF-36 (36 items), or Multiple Sclerosis Quality of Life-54 (MSQOL-54) (a combination of SF-36 and 18 MS-specific questions)—that they found to summarize most accurately their quality of life. Two percent preferred EQ-5D, 5% preferred SF-36, 17% preferred the markedly longer MSQOL-54, and the remaining 76% preferred a combination of selected items from the three. About 90% of patients found all three instruments acceptable and easy to respond to.

Nilsson et al. [40] studied patient satisfaction (n = 463) with the SF-36 and EQ-5D and patients’ perspectives regarding health outcome assessment within a routine health-care context. Patients completed the SF-36 and the EQ-5D before and after ordinary interventions. An evaluation form was also used to capture respondent satisfaction, where response alternatives for the cognitive response process ranged from “very easy” to “very hard” and patient-perceived content validity ranged from “very good” to “very
and questionnaire length (P length and response rate). Meta-analysis of the association between questionnaire length of patients found the SF-36 comprehensible, whereas 75% re-
sponded to were 54% for the SF-36 and 60% for the EQ-5D, respectively (P = 0.001). Patients considered both questionnaires to be equally applicable (68%). Even though the SF-36 consists of 36 items compared to 5 for the EQ-5D, it was preferred for routine assessment by 25% of patients, whereas only 8% favored the relatively shorter questionnaire (P = 0.001). The authors concluded that questionnaire length and ease of response were not crucial factors influencing the choice between the SF-36 and EQ-5D.

Discussion

Respondent burden, as measured by response rate, is frequently used as an incentive to develop brief instruments and abbreviate existing ones and is mentioned as a general obstacle in clinical practice to the administration of PROs. The aim of this literature review was to compile and evaluate systematically evidence of a potential relationship between questionnaire length and respondent burden. In total, 25 studies were found that examined the relationship indirectly by means of response rates; of these, only six reported significantly reduced response rates when longer questionnaires were administered. In our meta-analysis, there was a greater chance of response when patients were presented with a comparatively shorter questionnaire. Results, however, of the test of homogeneity indicated that factors other than length may have been at least as important per se. We found only three studies examining respondent burden more directly.

Measuring burden

Some studies have used time to complete a questionnaire as a measure of burden, for example, that of Haley et al. [41]. Because relying on time alone assumes that completing a questionnaire is burdensome by itself and that the passage of more time equals more burden without testing the presumption against any mea-
sure, these studies were not included in the literature review. Time per se provides no information as to whether the time spent on the questionnaire was burdensome. Comparatively shorter time to complete a questionnaire could even reflect a reduced effort to engage in the questions asked. Consequently, longer time to complete a questionnaire could imply that the respondent was more motivated and hence less bur-
dened. Most studies investigating respondent burden have done so by hypothesizing that higher response rates imply reduced bur-
den. Although this hypothesis is plausible, it is at best an assum-
ption because response rates do not provide any information as to why a questionnaire is not returned. Studies involving some type of patient input would help to clarify the reasons for not returning a questionnaire. Our systematic search found only three studies that touched on the concept of respondent burden by directly asking the patients which questionnaire they preferred and why.

Measurement context

Edwards et al. [23] previously reviewed factors relating to response rates of postal surveys, including published and unpublished arti-
cles spanning decades from various fields ranging from medicine to marketing. It was reported that a response was more likely when short questionnaires were used. Because the result of the test for homogeneity of the OR, however, was highly significant (P ≤ 0.00001), other factors are likely to have influenced the results. Also, there is reason to believe that a respondent’s motivation to complete a marketing survey would differ from a scenario in which the outcome of the assessment is believed to be important for evaluation of their health status. In our study, the findings are presented based on populations with and without clinical affilia-
tion, and similar motivational differences may apply to the study of Edwards et al. Typically, epidemiological studies rely on regist-
tries with patients who may have no affiliation to the organization conducting the research. However, the subject of an epidemiolog-
ical survey may be of greater relevance to the respondent than that of most marketing studies. Not only are studies with a clinical foundation based on a topic of potentially crucial relevance to affiliated respondents, but the patients involved may also believe that returning the questionnaire may affect their treatment out-
come. A concern that has frequently been raised is that response burden may be greater in severely ill patients, older individuals, children, and some demographic groups [4]. This concern may,

Results

Meta-analysis of the association between questionnaire length and response rate

In the meta-analysis, a general association between response rate and questionnaire length (P ≤ 0.0001) was found; response rates were somewhat lower for longer questionnaires. The Cochran-
Mantel-Haenszel estimate of the common OR over all strata was 1.14 and the 95% CI was 1.08–1.21. The P value for the Breslow-Day test was 0.03, indicating heterogeneity in ORs by study. Hence, the validity of the estimate of the common OR can be questioned, and this value should be interpreted with caution. Consequently, there is no clear indication that the difference in overall response rates is attributable to the differences in length between the questionnaires. Data on the numbers of pages used for shorter and longer questionnaires are not included in the analysis but are illustrated in Figure 1.
however, be overestimated and needs to be verified in the population of interest because several studies have failed to find any consistent relationship between compliance and population characteristics [42–46]. For example, 51 of 68 (75%) terminally ill patients with a life expectancy of less than 2 months reported no burden associated with completing an extensive battery of questionnaires [47].

**Psychometric and statistical issues**

A general concern with the shortening of instruments is validity and reliability. Instruments relying on one single question have, in general, very poor psychometric properties. For example, reduction in the number of items may result in a limited scope, which could lead to an instrument being insensitive to changes. A more specific issue is the number of studies needed to detect a significant difference in response rate between shorter and longer questionnaires. In this study, 6 of 25 studies reported lower response rates for participants who received a longer questionnaire. However, a marginal yet statistically significant difference was found when data for all 24,326 participants were included in an analysis stratified by study. When missing data are the only measure of respondent burden, the kind of missing data is important. Are the data missing at random, completely at random, or not at random? Reduced response rate and missing data will lead to a smaller sample size that, in turn, will reduce the probability of meeting study objectives. Only if the data are missing completely at random can this be mitigated by sample size calculations before starting a study. Of more concern is the potential bias caused by missing data that cannot be considered missing completely at random. Eleven of the 25 studies that used response rates as a measure of burden included information regarding missing data and partial responses. As methods for reporting partially missing data varied, no comparative analyses were performed. Overall, seven studies reported that there were no differences in data quality, one reported better data quality for the comparatively longer questionnaire and three for the shorter questionnaire. Finally, it should be noted that in a number of the cited studies, the reported response rates were so low that the generalizability of the findings was compromised; for example, Cunningham et al. [20] reported that only 22% of the patients returned the longer questionnaire. There is some evidence indicating that mode of delivery influences response rates (i.e., home interviewing and personal delivery is preferable to postal delivery [15,48]), and it has been proposed that the method by which a questionnaire is distributed is of greater importance than its length [15].

**Comparisons of length or content?**

An inherent problem with comparisons of questionnaire length is the relativity of the object of study. In the studies discussed, the number of pages compared was the most commonly reported measure. The meaning of “short” varied from 1 item to 16 pages, whereas “long” could be anything from 2 to 36 pages. Not only were there differences between studies, but also within studies. Although the increment expressed as a percentage for a study comparing one with two pages is equal to that of one comparing 16 with 36 pages, the additional time spent completing the questionnaire is clearly greater in the latter case. Thus, the number of pages may not be the most ideal measure of the effort needed to complete a questionnaire. Furthermore, the items comprising an instrument may differ vastly in complexity, for example, in terms of number of response options. The response burden is likely to be more pronounced when a respondent completes a shorter questionnaire with complex response choices compared to a longer questionnaire with more straightforward response alternatives. Another complication is that in most cases, it is impossible to separate out the effect of increased length from the content that is being added or removed when comparing two versions of the same questionnaire. This notion is substantiated by the significant result of the homogeneity test. It should also be noted that not all the studies reviewed compared different versions of the same questionnaire—some included comparisons of different questionnaires. For instance, Hassan and Weymuller [38] compared the generic SIP to the disease-specific and shorter UW-QOL. Questionnaires are often abridged when they are perceived to be too lengthy. If an instrument is successfully abbreviated from a content perspective, the reason that response rates would be higher for the shorter alternative is as likely to be related to content as to length—the least relevant items may have been the ones discarded. However, abridged and designated short instruments are rarely developed on the basis of patient input, but rather are guided by the outcome of a psychometric validation (e.g., factor analysis). Because length and content are difficult to separate out, the subject of quantity is indiscernible from that of quality. Thus, content validity, the degree to which the instrument measures the intended idea, is of equal importance to the length of the questionnaire per se. Evidence of content validity may be obtained from qualitative studies supporting the appropriateness of selection of domains and subordinate items of an instrument in relation to intention, target population, and utility [1]. Furthermore, the content validity of an instrument is dependent on the population studied, disease, and treatment. None of the reviewed studies that used response rates as outcome to study influence of length on response burden investigated whether items, domains, and scoring were found to be comprehensive and relevant from a patient’s point of view. Evidence of other types of validity (e.g., construct validity) will not diminish the need for content validation because the instrument needs to demonstrate that it measures all relevant aspects [1]. Three studies were identified that relied on patient input regarding questionnaire length. In two of these, patients were simply asked which instrument they preferred [38,39]. Findings from these studies indicate that patients would prefer a disease-specific instrument over a generic one, even when the instrument is more voluminous. A similar conclusion could be derived from the study conducted by Nilsson et al. [40], which both asked for the preference of instrument and included more fine-grained response alternatives for the recording of preference and comprehensibility. These studies exemplify how patient input may be obtained when deciding which instrument to use. Shifting focus from the length of the questionnaire to the content is likely to affect response rates positively; if the questions are deemed relevant, patients are more likely to be motivated to respond.

Concern about respondent burden has mostly focused on reducing the number of items in a questionnaire, but respondent burden may be more profound when the patient is asked to complete multiple questionnaires [49,50]. For example, the burden may be greater when the questionnaires included in a study reflect similar concepts (e.g., when clinical trials involve multiple questionnaires being administered on numerous occasions covering similar or identical concepts for research purposes). Rather than including similar shorter questionnaires, reducing the overlap of content between instruments may be more beneficial with regard to reduction of response burden. Also, given increased regulatory demands on questionnaires used in clinical trials [1], which may result in increased sampling density, there is a risk that repeated assessments may strain the patient. However, knowledge regarding the association of response burden with repeated assessments is limited, and future studies should investigate the potential relationship between sampling density and degree of response burden.
Introduction

Given the weak support for an association between questionnaire length and response burden, decisions on the choice of instrument are best based on the quality of the content from the patient’s point of view rather than the length per se.

Source of financial support: This study was financed by AstraZeneca R&D, Mölndal, Sweden.

References


Conclusion

Given the weak support for an association between questionnaire length and response burden, decisions on the choice of instrument are best based on the quality of the content from the patient’s point of view rather than the length per se.

Source of financial support: This study was financed by AstraZeneca R&D, Mölndal, Sweden.