5 Making the case for 'good enough' rape-prevalence estimates

Insights from a school-based survey experiment among Norwegian youths

Kari Stefansen, Mette Løvgren and Lars Roar Frøyland

Introduction

How to measure rape precisely has been a key topic in feminist research on violence against women since the 1980s. The issue of rape estimates is currently high on the agenda among jurisdictions that seek to develop prevention policies towards rape and other forms of sexual violence – such as Norway, the empirical context of this chapter. In national policy documents, prevalence estimates are held to be important because they shed light on the 'dark figure' of rape, i.e. the gap between incidents reported to the police and the real scope of the problem (e.g. Ministry of Justice and Public Security, 2012). States' responsibilities to monitor the prevalence of rape are also incorporated in supranational policies, such as the 2011 Council of Europe Convention on preventing and combating violence against women and domestic violence (known as the Istanbul Convention), which all the Nordic countries have ratified.

In Norway, the study of rape prevalence is still in the early stages. In 2008, when the first governmental committee was appointed to outline new directions for rape prevention (Ministry of Justice and Public Security, 2008), no nationally representative surveys were available. The committee therefore estimated the national prevalence based on a survey conducted in Oslo, the capital city (Pape & Stefansen, 2004). The prevalence from Oslo is not necessarily representative for the whole country, however, due to the city's significantly higher proportion of inhabitants from migration backgrounds and a larger dispersion of economic and cultural resources than is the case among the general Norwegian population.

Since 2010, two studies have been conducted on nationally representative samples of the adult population (Steine, Milde, Bjorvatn, Grønli, & Nordhus, 2012; Thoresen & Hjemdal, 2014). Both departed from the legal definition of rape and used behaviourally specific questions to measure the prevalence of acts that fall within this definition. Norwegian law defines rape as an act of penetration by the use of force or threats, as well as sexual assault against a person who is unable to consent (penal code section 191; see Skilbrei & Stefansen, 2018). The definition further includes incidents where direct contact has occurred between the sex organ

or finger of the accused (or an object or someone acting on the accused's behalf) and the victim's vagina or anus, as well as the insertion of a penis into the victim's mouth (Haugen & Efjestad, 2015, p. 344). The code also extends to incidents involving masturbation. The threshold for what counts as force is generally considered very low and is in line with legislation elsewhere with a focus on non-consent (Ot. prp. nr. 22 [2008–2009]).

The two studies mentioned above had different samples and used different methodologies, but they reported similar rates of rape prevalence. One study (Steine et al., 2012) used a six-item measure that included incapacitated sexual assault and the experience of feeling pressured to have intercourse without the use of threats or force, as well as forced or threatened penetration of bodily orifices by a sex organ, finger or object. In comparison, Thoresen and Hjemdal (2014) used a four-item measure covering different forms of penetration by the use of force or threats (against the victim or someone close to the victim), and they did not ask about pressure (which in most cases would not qualify as rape in the legal sense). They also excluded an item on incapacitated sexual assault. Steine et al. (2012) reported a total prevalence of rape experiences after the age of 16 to be 16.3% for women and 1.6% for men. The item on pressure revealed the highest number by far for women: 11.6% had experienced being pressured to have intercourse. For the other items, the numbers varied from 1.6% to 4.1% (see Steine et al., 2012, Table 5.3, p. 954). Thoresen and Hjemdal (2014) reported somewhat lower lifetime rates for rape: 9.4% for women and 1.1% for men.

Although very similar in their results, the two studies illustrate a general point: that to measure rape is also to construct rape. What we include and what we exclude speaks to and contributes to specific understandings of a given phenomenon. While this might be a rather banal insight, it is a fundamental one if we are to construct reliable measures of the prevalence of rape across time and context.

Given this background, the aim of this chapter is twofold: first, to introduce ongoing discussions about rape measurement in the international literature, and second, to present and discuss results from a large-scale school-based survey among Norwegian youth that featured an experiment in which two different rape measures were presented to the participants. We ask: What can we learn from comparing the two measures, and how can we assess the quality of the 'better' measure? In what sense is it 'good enough' as a prevalence measure?

Rape measurement: the international debate

The development of methodology for the measurement of rape is intimately connected to the 1980s work of the US psychologist Mary P. Koss in developing the Sexual Experiences Survey (SES), a set of behaviourally specific questions about rape and other forms of sexual assault that were first published in Koss and Oros (1982) and later revised in Koss, Abbey, Campbell, Cook, and Norris (2007). The original aim of the SES was to capture all acts that had been criminalised as rape. It therefore included both less violent attacks and

rape in intimate relationships, such as date rape and marital rape. During the eighties and nineties, studies based on the SES found rape and other forms of sexual assault to be much more prevalent than estimates from general crime surveys and police statistics had indicated (Koss, 1992, 1993; Koss, Gidycz, & Wisniewski, 1987). These studies were also the subject of criticism, however. Gilbert (1993), for example, argued that their definition of rape was too broad and that the questions that were posed captured incidents that did not meet the legal criteria for rape. In essence, Gilbert critiqued Koss and the SES for casting the net too wide.

In a comprehensive report about rape measurement, Fisher and Cullen (2000) described the situation at the end of the 1990s as being characterised by controversy. Their report, which reviewed research to date on rape measurement and aimed to outline directions for future research, then sparked new debates. Fisher and Cullen followed Koss in the rejection of stand-alone questions that ask directly about exposure to rape, forced intercourse or similar formulations. Such questions cast the net too narrowly and have poor validity, because most respondents will respond in accordance with stereotypical ideas about rape. The result, most often, is massive underreporting of rape, especially of rapes committed by partners and acquaintances and of date rape (Kahn, 2004; Kahn, Jackson, Kully, Badger, & Halvorsen, 2003; Stefansen & Smette, 2006).

Fisher and Cullen (2000) acknowledged Koss and colleagues' important contributions to rape measurement, especially the shift from general to behaviourally specific questions. The latter approach offers a way around cultural beliefs about rape and rapists because it focusses on specific acts – which researchers can then calculate into rape measures – although they also criticised the one-step model that the SES represents. Fisher and Cullen's reservation was that not all the recorded incidents would likely meet the legal criteria of rape had the researchers followed up with more detailed questions. Hence, the one step-model may lead to exaggerated estimates of rape. As the authors argue,

> the risk of this one-step approach – that is, of not having an incident report – is that it may include reports of victimization that, on closer scrutiny, would not qualify legally as a rape or other types of sexual victimization.
>
> (Fisher & Cullen, 2000, p. 324)

They instead argue that a two-step model would yield more accurate estimates. In this model, the respondents are first presented with a list of screening questions about different forms of sexual assault that are intended to trigger the respondents' memory about relevant incidents. Respondents who answer yes to one or more of the first set of questions are then asked a new set of questions related to each incident. Prevalence estimates are based on the information from the second step and will only include incidents that meet the legal criteria for rape. According to Fisher (2009), this model ensures both that incidents that are mistakenly included in the material as rape through the first step can be excluded and that other forms of sexual assault can be reclassified as rape through the second

step and subsequently used in the prevalence estimate. Hence, the model solves two problems: it avoids both casting the net too wide (by excluding 'false' rapes) and casting it too narrowly (by including 'false' non-rapes).

Fisher (2009) has further shown that asking more questions and using behaviourally descriptive questions in the first step includes a higher number of respondents in the incident report and yields a higher percentage of rape victims in the second step. Koss and colleagues (in Cook, Gidycz, Koss, & Murphy, 2011) have responded to Fisher and Cullen's (2000) critique and the two-step model's potential for more accurate measurements. They contend that it is ethically problematic to exclude respondents who have indicated that they have been raped, even if they respond differently later in the questionnaire. They refer to the two-step method Fisher (2009) used as 'confrontational' (p. 210) because the interviewer asks the respondent about what 'actually' happened during the reported incident. This critique is a reminder that concerns other than accuracy – such as the potential for psychological distress – must be taken into account when assessing the usefulness of rape-measurement models.

A pragmatic approach to rape measurement

While these authors disagree on the merits of the two-step model compared to the one-step model, what they have in common is a commitment to work towards establishing best practice for rape-measurement research. These methodological debates may give the impression that it is possible to design an instrument that will capture all incidents of rape in a surveyed sample. Fisher and Cullen, however, comments that researchers now acknowledge that it is unlikely that any model can avoid all sources of error:

> In particular, researchers have come to realize that conceptually defining and then operationalizing sexual victimization are complicated and, to a degree, imperfect enterprises – especially when deciding when an unwanted sexual advance crosses the line from imprudence to criminal behaviour.
>
> (Fisher & Cullen, 2000, p. 320)

Fisher and Cullen (2000) also stress that the issues of definitions and operationalisations of rape and other forms of sexual violence are epistemological, since 'they raise critical issues about what we know and how we know it' (p. 340).

In a review article, Krebs (2014) summarised the debate on rape measures and argued for a pragmatic approach. The key issue, according to the author, is to understand when different models are suitable and how they can be improved. In his view, it is more fruitful to establish consensus on areas that different models have in common rather than to continue debating which model is best. This consensus would include the fact that instruments that use behaviourally specific questions within self-administrating surveys seem to work well.

His approach, which we follow here, represents a call for establishing consensus on 'good enough' instruments. In our opinion, doing so does not necessarily conflict with a quest for better or ideal instruments for population estimates of rape but represents a complementary approach. In the next section, we describe the study that our analyses are based on.

The study

Sample and participants

Our data are drawn from a school-based study among students aged 18–20 years in Norway in the last year of upper secondary school. The data were collected in 2015. The students answered a comprehensive electronic questionnaire about different types of victimisation, including questions about sexual assault and rape. The study was a follow-up to a similar study conducted in 2007 (Mossige & Stefansen, 2007) and therefore used the same sample of schools. For the 2007 study, a representative sample of schools was provided by Statistics Norway following a two-stage procedure. The schools were first allocated to five geographical regions to ensure participation from the whole country and then stratified within each region into three categories: vocational track only, academic track only and combined. The 67 participating schools were then selected to ensure proportional allocation. These schools were contacted again and asked to participate in the 2015 study. A total of 41 schools from the 2007 study participated; in addition, eight replacement schools were selected to fit the relevant strata, yielding a total sample of 49 schools. From these schools, 4,530 students participated, resulting in a response rate of 66.2%. Schools and groups within schools with a response rate below 10% were excluded from the material. In these cases, the survey had not been administered according to instructions.

The students completed the survey during two consecutive school sessions with a teacher present in the classroom. The mean age of the respondents was 18.4 years; 1,766 respondents were male, and 2,580 were female. The gender difference in the sample was caused by a larger proportion of academic-track schools (in which girls are overrepresented). The gender was missing for 65 respondents. The analyses discussed in this chapter are restricted to students between 18 and 20 years old with non-missing responses to the question of gender. For the rape items, we included students with non-missing responses on at least one item of the multi-item measure (described below). The final sample for the analyses presented in this chapter consisted of 4,346 participants.

Because we recruited respondents through schools, dropouts were not part of the sample, which may have affected the results – those who drop out differ from those who do not in terms of risk and protective factors associated with rape (Mossige & Stefansen, 2007, 2016).

Questions and variables

The items about rape were part of a survey instrument in which a range of sexual-victimisation experiences were mapped, including unwanted touching and attempted rape. Those items that were designed to capture experiences with rape followed items about other forms of sexual victimisation. The latter items could therefore work as memory cues for unwanted sexual experiences, including rape.

Using a survey-experiment design, we randomly assigned the respondents to different versions of the instrument: one with a single question about rape and the other with five behaviourally specific items. In the questionnaire, both versions had the following introductory phrase: *Before/after the age of 13, have you been subjected to any of the following against your will?* Alternative 1 read: 'You have been raped'. Alternative 2 consisted of the following items: someone 'forced or threatened you to have intercourse', 'forced or threatened you to have oral sex', 'forced or threatened you to have anal sex', 'injected fingers or objects into your vagina or anus', and 'has had sex with you while you were asleep or too drunk to object'. The answer options for both versions were 'No, never', 'Yes, once', and 'Yes, several times'. As expected, the randomisation achieved balance across variables such as gender, age, parents' socioeconomic status and school track (academic or vocational). Hence, all observed differences in reported prevalence rates are attributable to being asked either broad or behaviourally specific questions (cf. Løvgren, Stefansen, Smette, & Mossige, 2017).

In the following, we investigate how the behaviourally specific questions compare to the broad one-item question. First, we compare the responses from girls and boys; we then compare the responses between girls within different subgroups to test if different wordings had greater or lesser impact within certain groups compared to others. This part of the study is exploratory, meaning that we do not have existing research on which to form hypotheses on the impact of different wording between subgroups, although we do expect that the behaviourally specific instrument will yield significantly higher prevalence rates compared to the one-item question, and that this result will hold for both genders and across different subsamples. Because the sample size was reduced when we analysed subgroups, the lower prevalence rates of rape experiences among boys compared to girls led to problems with statistical power; therefore, these analyses were only conducted for girls.

We do not distinguish between experiences before and after the age of 13. Most of the reported experiences of rape took place after the age of 13 (analysis not shown). We used dummy variables in all analyses. For the one-item question (labelled 'Rape1' in this chapter), a value of 1 was given to those who reported having been raped at least once, either before or after the age of 13. We constructed two different dummy variables based on the five-item alternative, with a value of 1 indicating at least one positive answer. The 'Rape4' measure consists of the four items that specify different forms of penetration

with the use of force or threats, excluding the item on incapacitated sexual assault. The 'Rape5' measure includes all five items.

We had two reasons for constructing two variables rather than one. First, we wanted to compare our results with those of another Norwegian study in which incapacitated sexual assault was not included in the reporting of the results (Myhre, Thoresen, & Hjemdal, 2015). Second, the 'incapacitated sexual assault' item used the general term 'sex' rather than specifying an act of penetration. We address possible implications of this choice of wording in the discussion section below.

In addition to the questions about rape, we use variables on family economy, parents' work-life participation and choice of academic track in upper secondary school in the analyses. Family economy is measured by a single question: 'Over the last two years, has the economic situation of your family been good or bad?' The response options were 'Good all of the time', 'Mostly good', 'Neither good nor bad', 'Mostly bad', and 'Bad all of the time'. The measure was dichotomised into two dummy variables: *good economy* ('good all of the time' and 'mostly good'), and *poor economy* ('mostly bad' and 'bad all of the time'). Two variables measure whether the respondents' fathers and mothers were currently working (either full time or part time), and two variables measure if either parent received social benefits. Finally, a single question measures whether the respondents followed the academic or vocational track in upper secondary school. All these variables were coded into dummy variables (1 for affirmative answers; otherwise 0).

We tested for mean differences across groups using *t*-testing. Because we have tested multiple hypotheses and have thus increased the chances of test results displaying as significant when they are in fact not, we advise caution in interpreting the results. Specifically, with a significance level of $p = 0.05$, one out of every 20 analyses could, out of statistical necessity, be expected to return a significant result, more commonly known as a false positive finding (type I error).

Results

As demonstrated in Table 5.1, very few male students reported being subjected to sexual assault, by either measure. The absolute numbers as well as the rates are higher among female students. The differences between males and females are significant for all items except the one on anal penetration. Both for boys and girls, incapacitated sexual assault is the more common experience. Among girls, being subjected to enforced penetration of a finger or object and by a penis are more common than enforced oral and anal penetration. Among boys, about 1% reported all the different forms of penetration.

Interestingly, among girls, the one-item question about rape and the item on forced or threatened intercourse had very similar rates, only differing by half a percentage point. This finding could suggest that what the subsample which received the one-item measure had reported as rape were incidents of forced or

Table 5.1 All items. Prevalence of positive responses. By gender. Frequency and percent.

	Boys			*Girls*			
	Percent	*Freq.*	*N*	*Percent*	*Freq.*	*N*	*T-test p-value*
Broad question							
You have been raped	1.1	9	832	5	62	1259	<0.001
Behaviourally specific questions							
Forced or threatened you to have intercourse	0.9	8	872	4.5	57	1275	<0.001
Forced or threatened you to have oral sex	0.9	8	873	2.9	37	1271	<0.01
Forced or threatened you to have anal sex	0.8	7	870	1.7	21	1271	>0.05
Injected fingers or objects into your vagina or anus	0.9	8	873	6	76	1273	<0.001
Someone has had sex with you while you were asleep or too drunk to object	2.2	19	868	7.6	96	1270	<0.001

threatened intercourse. This situation would correspond with the common finding that assaults involving intimidation or the explicit use of physical power more often are considered rape (Ryan, 2011; Stefansen & Smette, 2006). Because the items were distributed to different subsamples and not the same respondents, we cannot test the degree of overlap.

How do the different measures of rape compare? Table 5.2 shows the results for the dummy variables Rape1, Rape4 and Rape5. Because each variable is a dummy, the mean reads as a percentage; a mean of 0.05 means a 5% prevalence.

Table 5.2 demonstrates – as expected – that the measures based on behaviourally specific questions elicited higher rates of sexual assault among both boys ($p < 0.001$) and girls ($p < 0.001$). Further, the gender differences are notable. For girls, the prevalence rate increases from 5% to 14% from the one-item to the five-item measure, and for boys, the rate increases from 1% to 3%. Table 5.2 also demonstrates that the difference in victimisation rates between boys and girls increases as more items are included in the measure. The difference is four percentage points for the one-item question, seven percentage points for the four-item measure and eleven percentage points for the five-item measure.

The analyses so far relate to the whole sample. We also investigated if the measures based on the behaviourally specific questions would increase the prevalence of rape to a similar degree within different subgroups of *girls*. This would strengthen the evidence for the instrument if so. For these analyses, we

Table 5.2 Three measures of rape. Boys and girls. Between-group mean comparison. *t*-Test.

	Boys			*Girls*			
	N	*Mean*	*Std. Dev.*	*N*	*Mean*	*Std. Dev*	t-*Test,* p-*value*
Rape1	832	0.01	0.1	1259	0.05	0.22	<0.001
Rape4	872	0.02	0.12	1276	0.09	0.29	<0.001
Rape5	873	0.03	0.17	1276	0.14	0.34	<0.001

chose a set of sociodemographic variables known to affect levels of sexual assault. Research in Norway has consistently shown that the risk of sexual victimisation is significantly higher among those with low socioeconomic status compared to those with high socioeconomic status (Mossige & Stefansen, 2007, 2016; Pape & Stefansen, 2004). The same pattern has been documented in international research (e.g. Phipps, 2009). For this study, we use indicators related to family economy, parents' labour-market status and choice of academic track in upper secondary school.

In the following analyses, we focus only on differences between two measures (Rape1 and Rape4) because we are uncertain about the types of experiences reported in the fifth item, on incapacitated sexual assault. The use of the general concept 'sex' in this item means that it is unclear if the reported incidents involved acts that are criminalised as rape in the Norwegian penal code, or if they represent cases of sexual assault that do not involve penetration (or penetration-like acts). We draw on Myhre et al. (2015), who found that around 50% of reported incidents on a similar item did not involve penetration. In order to ensure internal validity, we therefore proceeded with the four-item measure, given that all the acts covered align with the legal definition of rape (Table 5.3).

The group differences in rape prevalence are large but as expected. For example, the rate of rape is considerably lower among girls with a good family economy in the last two years compared to girls with a poor family economy in the same period. For the Rape1 measure, the prevalence is 4% for the former group and 12% for the latter group; for Rape4, the respective numbers are 8% and 19%. Not surprisingly, we see a nearly identical pattern for girls with parents who work or receive social benefits (variables that are indicative of the economic conditions of the family). We see corresponding but lower differences between the other groups. The difference for Rape1 between girls in academic and non-academic tracks is close to three percentage points, and for Rape4, the difference is around nine percentage points.

For most of the groups included in the study, the prevalence of rape was significantly higher when we used the four-item measure compared to the single-item measure. The non-significant result for having a poor family

Table 5.3 Rape measures and socio-economic characteristics. Girls. Between-group mean comparison. *t*-Test.

	Rape1			*Rape4*			
	N	*Mean*	*St. dev.*	*N*	*Mean*	*St. dev.*	*T-test*
Poor family economy	67	0.12	0.33	58	0.19	0.4	n.s.
Good family economy	925	0.04	0.2	953	0.08	0.27	***
Father working	1058	0.043	0.20	1093	0.09	0.28	***
Father social benefits	64	0.13	0.33	58	0.14	0.35	n.s.
Mother working	1059	0.046	0.21	1081	0.088	0.28	***
Mother social benefits	91	0.044	0.21	98	0.13	0.34	*
Academic track	892	0.043	0.2	940	0.076	0.26	**
Non-academic track	94	0.043	0.2	78	0.13	0.34	*

economy is most likely ascribable to the low *n*, as the *t*-test is a function of the magnitude of the observed differences and the sample size. In larger groups (e.g., the non-academic-track group), similar differences returned significant results at the 0.05 level. In general, the Rape4 measure resulted in approximately twice the rate of rape compared to the Rape1 measure. Following an academic track resulted in a somewhat smaller discrepancy between the two measures. But logistic regression analyses (not shown), with an interaction term between the different background characteristics and the given rape instrument, showed no significant results. As such, the relative difference in the proportion of rape measured by the two rape instruments does not differ significantly according to background characteristics.

Overall, the measures provide comparable differences in rape rates for all the groups we included, even though the total prevalence varied greatly according to the given instrument and among the different groups. We may thus conclude that the four-item measure works as intended. It captures a significantly higher number of rape experiences across different groups of girls compared to the one-item measure, thereby reducing underreporting. The measure hence provides an estimate of rape that is more in accordance with the penal definition of rape. Does this mean that it can be considered 'good enough'? This would depend on the criteria for defining what is good enough – to be discussed next.

Discussion

We have shown that the four-item measure works as intended and performs well across different groups. The measure, which is based on recommendations in the international literature, is part of a broader instrument that also surveys other forms of sexual assault (which then triggers participants' memory of unwanted sexual experience), and the items are behaviourally descriptive. The

items also cover acts that have been criminalised as rape in the Norwegian penal code. We would add that the measure has important advantages for examining the population in question (students in upper secondary school) and for the chosen methodology (i.e. a more general victimisation survey administered at school).

First, the measure is relatively short, meaning that it can be integrated into surveys on a wider range of topics, including other aspects related to sexual assault such as background characteristics, risk factors and health outcomes. The use of a short instrument also prevents exhaustion on the part of the respondents and subsequently reduces the number of missing answers on later items in the instrument. We found this to be the case in the present study, where 98% of those who had answered the first question in the instrument also answered the remaining questions. While the length of questionnaires and instruments is important for all samples, it is likely that the length is even more critical when conducting surveys among young people. A related point is that asking less detailed and probing questions may be less emotionally draining for students who have experienced rape – and therefore would be more suitable for use in classroom settings with other students and a teacher present, where an emotional reaction would be noticed.

Second, the items themselves are short and entail few elements; the aim of this approach is to limit the possibility for misunderstandings or differing interpretations. In formulating the items, we were inspired by Thoresen and Hjemdal's (2014) study, although the phrasing used in their study was much more complex. To exemplify, while we asked, 'Has someone forced or threatened you to have intercourse?', they asked, 'Has someone at any time forced you to have intercourse by the use of physical power or by threatening to harm you or someone you have a close relationship with?' While the first item is related to force or threats and the nature of the sexual act, the latter also specifies that the timeframe is unlimited, and the type of force and threats.

These advantages aside, what are the relevant criteria to assess the instrument as being 'good enough' for the production of rape estimates in a youth population?

As a first step, we would look at the rates of rape produced by this and similar instruments in different studies. Studies that produce similar results would speak to the validity of the instrument. In the introduction to this chapter, we detailed the results from two previous studies conducted among the adult population in Norway. Both used sets of behaviourally specific questions, and the results were very similar to what we have found in our study. Based on our four-item measure, close to 9% of the girls had experienced rape; the corresponding number in the adult study in which a four-item measure was used (Thoresen & Hjemdal, 2014) was only slightly higher. If we compare our five-item measure with the adult study that included incapacitated sexual assault, then the numbers are also quite similar. Among the girls in our study, the prevalence rate was 14%; among the women in the adult study, the rate was around 16% (Steine et al., 2012). Given that adults have had more time to accumulate

rape experiences, these higher rates among adults are as expected. We would also expect that the difference would be quite low, given that young people are considerably more at risk than adults and older people (see Kruse, Strandmoen, & Skjørten, 2013). But our results differ markedly from the results in Myhre et al.'s (2015) study among 16- and 17-year olds, which used the same four-item measure as Thoresen and Hjemdal (2014). The prevalence of rape was as low as 3.5% among girls and 0.2% among boys. The two-year age span between the respondents in Thoresen and Hjemdal's (2014) study and our study complicates comparisons. These two years of a young person's life most likely see a disproportionate number of sexual assaults compared to previous years. We suspect, however, that the marked difference in part is caused by the composition of the sample in Myhre et al.'s (2015) study. They drew their sample from the telephone directory and reached only 10.4% of the original sample, from which 66.2% chose to participate (p. 50), thus making it unlikely that the final sample was representative of the youth population. In addition to the low response rate, the study was not based on a self-completion questionnaire but on telephone interviewing, which involves personal contact between the participant and interviewer. The choice of methods is known to affect people's willingness to disclose sensitive issues, with self-administered methods yielding higher rates compared to methods that involve personal contact, either face-to-face or on the telephone (Krumpal, 2013).

A second – and perhaps the most important – step would be to evaluate the degree to which the measure covers the legal definition of rape. We have already commented on why we have excluded the item on involuntary incapacitated sex from the final analysis. Given the high percentage among both girls and boys who reported involuntary incapacitated sex in our study, we would expect that the inclusion of incapacitated rape (penetration or similar) would result in higher prevalence rates for both genders. How an item on incapacitated rape should be phrased is not a straightforward matter. The crime of incapacitated rape covers all forms of penetration by different means – and hence captures a broader set of acts than incapacitated penal–vaginal penetration. It is clear that our four-item measure is too narrow to capture all acts of incapacitated rape. By including the involuntary incapacitated-sex item, we would solve this problem but would at the same time capture a wider set of acts, thus leading to an overly high rape estimate. The solution for future research is to reformulate this item in a manner that would accord with the legal provision on incapacitated rape, or to apply the follow-up method that Thoresen and Hjemdal (2014) used and include confirmed incidents of incapacitated rape in the rape estimate.

What about the broad understanding of rape in the penal code? Rape in the penal code is not limited to 'completed' penetration but also covers incidents of direct contact between sex organs or sex organs and fingers; it also includes masturbation by force or threats of the accused as well as of the victim. The items on rape in our study are all focussed on incidents of rape in a more narrow sense – completed penetration – and hence set the bar

higher than the legal code does. Also, these items do not include acts of masturbation by the use of force or threat. How the rape code is to be interpreted is not spelled out in the code itself but is instead to be found in preparatory works and jurisprudence; hence, current law cannot be read directly from the letter of the law. Our instrument is flawed, in the sense that it is an operationalisation of the concrete paragraph on rape and not of current law in action. On the other hand, one could argue that the items we have included cover the main acts that legislators would wish to criminalise (i.e. that they are operationalisations of legislators' intent, so to speak), given, of course, that legislators intend to capture acts of a certain degree of seriousness that would represent more invasive intrusions than unwanted touching and the like. The more general point is that how we should interpret the general notion that population studies should depart from the penal definition of rape is not particularly clear (cf. Fisher & Cullen, 2000).

Another issue related to operationalisation concerns gender. In line with previous research, our behaviourally specific measures document a clear gendered pattern in the exposure to rape. These measures were intended to capture all acts considered rape under the Norwegian penal code. But the items we have included all address the experience of being penetrated, while the penal code also covers incidents of forcing someone to penetrate someone else, including the assailant.[1] In the literature, this type of sexual assault is labelled 'made to penetrate', 'forced to penetrate' and 'compelled penetration' (Hlavka, 2017; Stemple & Meyer, 2014; Weiss, 2010). None of the Norwegian prevalence studies have included questions about 'forced-to-penetrate' rapes, so the prevalence is unknown. Referring to studies from different countries and using different methods and questions, Weare (2018) concluded that 3–5% of the male population might have experienced rape in this sense.

We do not know how the inclusion of one or more items on 'forced-to-penetrate' incidents would have affected the results in our study and the gender difference in rape exposure we found, but such an inclusion would likely mean a higher prevalence among boys than we have reported in this study. The general point is that the use of rape measures that do not include those specific forms of rape that only men and boys experience contributes to constructing rape as the act of being penetrated, and therefore as a predominantly female experience. As for the evaluation of our measure, we would conclude that it is biased, and that the measure should be expanded with items that cover compelled penetration. For girls, however, the instrument most likely yields a more accurate estimate, even if it excludes certain incidents of incapacitated rape.

Conclusion

Our work in this chapter is inspired by Krebs (2014) and his call to work simultaneously to improve the one- and two-step approaches to the measurement of rape in population studies. We have detailed our experiences from using a one-

step model based on only five items in a school-based survey among youths. The aim of using these items was to capture rape as it is defined in the Norwegian penal code. Our analyses revealed that those measures that were based on behaviourally specific items yielded higher rates than the one-item general measure did, which held for both genders and for different subgroups of girls. Due to the low N, within-group differences could not be tested for boys – hence we do not know if the results are the same for boys. The possibility of including the measure in a school-based survey meant that the response rate was high and that the rape estimates were based on nationally representative data from youths (at least those who were still in school).

Our conclusion is that the measure in some senses is 'good enough'. For girls, it does capture most incidents of rape, according to the legal concept of rape; for boys, it is too narrow and needs to be reworked to include compelled penetration. As an operationalisation of current law on rape in Norway, the measure is probably too strict – given that it is based on the letter of the law and not on how it is interpreted in preparatory works and legal practice. We would caution against including more acts and less clear experiences, however, as the net then could be cast too wide and could include offences that are located in the grey area between rape and other forms of sexual violence.

We would also welcome critical debate of the idea that societies need exact numbers on rape prevalence. There is much to be learned from the good-enough measures used in studies that can provide representative data, especially from studies that monitor developments in rape prevalence over time. We would also caution against the idea that the higher the prevalence, the more trustworthy or useful the study: a high prevalence may be the result of asking overly general questions. In our study, the inclusion of an item about involuntary incapacitated *sex* (which is not necessarily rape in Norwegian law) had a dramatic effect, for instance, by doubling the percentage of affected boys.

On this note, we will make two final points. First, the one-item general approach has been heavily criticised in international scholarship because it leads to underreporting and a skewed description of who is affected by rape. This critique is well substantiated. The value of general questions is seldom explored or explicated, however. In our study, this question made for an interesting comparison with the behaviourally specific measures. We also believe that it can be very fruitful to follow developments over time in both the prevalence of self-defined rape and the difference between self-defined rape and rape as measured by behaviourally specific questions. One hypothesis concerning the latter would be that the difference will decrease over time because of the heightened attention to the issue of what counts as rape in prevention campaigns and public debate.

Second, the value of prevalence research for monitoring purposes and for prevention efforts needs to be discussed. Prevalence estimates need to be complemented by research that aims to understand the aetiology of rape on the population level. For example, one recent study (Heinskou, Friis, Ejbye-Ernst, & Liebst, 2017) has illustrated how prevalence studies can be used to identify how

rape is located in time, space and relations. In Heinskou et al.'s analyses, the authors examine exposure to rape in relation to a range of factors, including who the assailant was, when it happened and what the consequences of the incident were, both during the situation and afterwards. Such analyses may be developed further to monitor where different forms of rape are socially located and how they decrease or increase over time in different groups. Doing so would then pave the way for more ambitious theoretical work on the social patterns of rape in modern societies.

Note

1 Penal code section 291, letter c, defines an act to be rape if 'through violence or threatening conduct [someone] makes a person engage in sexual activity with another person, or perform[s] acts corresponding to sexual activity on himself/herself'. The code applies the same maximum penalty as committing forceful penetration or similar types as well as incapacitated rape. See https://lovdata.no/dokument/NLE/lov/2005-05-20-28/KAPITTEL_2#KAPITTEL_2.

References

Cook, S. L., Gidycz, C. A., Koss, M. P., & Murphy, M. (2011). Emerging issues in the measurement of rape victimization. *Violence against Women*, *17*(2), 201–218.

Council of Europe. (2011). The Council of Europe Convention on preventing and combating violence against women and domestic violence.www.coe.int/fr/web/conventions/full-list/-/conventions/rms/090000168008482e

Fisher, B. S. (2009). The effects of survey question wording on rape estimates: Evidence from a quasi-experimental design. *Violence against Women*, *15*(2), 133–147.

Fisher, B. S., & Cullen, F. T. (2000). Measuring the sexual victimization of women: Evolution, current controversies, and future research. *Criminal Justice*, *4*, 317–390.

Gilbert, N. (1993). Examining the facts: Advocacy research overstates the incidence of date and acquaintance rape. In Donileen R. Loseke, Richard J. Gelles & Mary M. Cavanaugh (Eds.), *Current Controversies on Family Violence* (pp. 120–132). Newbury Park, CA: Sage.

Haugen, F., & Efjestad, J. S. (2015). Strafferett – Håndbok (4th ed.). Oslo: Cappelen Damm Akademisk.

Heinskou, M. B., Schierff, L. M., Ejbye-Ernst, P., Friis, C., & Liebst, L. S. (2017). *Seksuelle krænkelser i Danmark: Omfang og karakter*. København: Aalborg Universitet, Københavns Universitet.

Hlavka, H. R. (2017). Speaking of stigma and the silence of shame: Young men and sexual victimization. *Men and Masculinities*, *20*(4), 482–505.

Kahn, A. S. (2004). 2003 Carolyn Sherif award address: What college women do and do not experience as rape. *Psychology of Women Quarterly*, *28*(1), 9–15.

Kahn, A. S., Jackson, J., Kully, C., Badger, K., & Halvorsen, J. (2003). Calling it rape: Differences in experiences of women who do or do not label their sexual assault as rape. *Psychology of Women Quarterly*, *27*(3), 233–242.

Koss, M. P. (1992). The underdetection of rape: Methodological choices influence incidence estimates. *Journal of Social Issues*, *48*(1), 61–75.

Koss, M. P. (1993). Detecting the scope of rape: A review of prevalence research methods. *Journal of Interpersonal Violence*, *8*(2), 198–222.

Koss, M. P., Abbey, A., Campbell, R., Cook, S., Norris, J., Testa, M., … White, J. (2007). Revising the SES: A collaborative process to improve assessment of sexual aggression and victimization. *Psychology of Women Quarterly*, *31*(4), 357–370.

Koss, M. P., Gidycz, C. A., & Wisniewski, N. (1987). The scope of rape: Incidence and prevalence of sexual aggression and victimization in a national sample of higher education students. *Journal of Consulting and Clinical Psychology*, *55*(2), 162.

Koss, M. P., & Oros, C. J. (1982). Sexual experiences survey: A research instrument investigating sexual aggression and victimization. *Journal of Consulting and Clinical Psychology*, *50*(3), 455.

Krebs, C. (2014). Measuring sexual victimization: On what fronts is the jury still out and do we need it to come in? *Trauma, Violence, & Abuse*, *15*(3), 170–180.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, *47*(4), 2025–2047.

Kruse, A. E., Strandmoen, J.-F., & Skjørten, K. (2013). *Menn som har begått voldtekt – En kunnskapsstatus*. Oslo: Nasjonalt kunnskapssenter om vold og traumatisk stress.

Løvgren, M., Stefansen, K., Smette, I., & Mossige, S. (2017). Barn og unges utsatthet for fysisk vold fra foreldre. Endringer i mild og grov vold fra 2007 til 2015. *Tidsskriftet Norges Barnevern*, *94*(2), 110–124.

Ministry of Justice and Public Security. (2008). *NOU 2008: 4 Fra ord til handling, Bekjempelse av voldtekt krever handling*. Oslo: Ministry of Justice and Public Security.

Ministry of Justice and Public Security. (2012). *Handlingsplan mot voldtekt 2012–2014 [Action plan against rape 2012–2014]*. Oslo: Ministry of Justice and Public Security.

Mossige, S., & Stefansen, K. (2007). *Vold og overgrep mot barn og unge: En selvrapporteringsstudie blant avgangselever i videregående skole*. Oslo: NOVA.

Mossige, S., & Stefansen, K. (2016). *Vold og overgrep mot barn og unge: Omfang og utviklingstrekk 2007–2015*. Oslo: NOVA.

Myhre, M., Thoresen, S., & Hjemdal, O. (2015). *Vold og voldtekt i oppveksten: En nasjonal intervjuundersøkelse av 16-og 17-åringer*. Oslo: NKVTS.

Pape, H., & Stefansen, K. (2004). *Den skjulte volden: En undersøkelse om Oslobefolkningens utsatthet for trusler, vold og seksuelle overgrep*. Oslo: NKVTS.

Phipps, A. (2009). Rape and respectability: Ideas about sexual violence and social class. *Sociology*, *43*(4), 667–683.

Ryan, K. M. (2011). The relationship between rape myths and sexual scripts: The social construction of rape. *Sex Roles*, *65*(11–12), 774–782.

Skilbrei, M.-L., & Stefansen, K. (2018). *Seksuell vold: En samfunnsvitenskapelig innføring*. Oslo: Cappelen Damm Akademisk.

Stefansen, K., & Smette, I. (2006). Det var ikke en voldtekt, mer et overgrep: Kvinners fortolkninger av seksuelle overgrepserfaringer. *Tidsskrift for Samfunnsforskning*, *47*(1), 33–56.

Steine, I. M., Milde, A. M., Bjorvatn, B., Grønli, J., Nordhus, I. H., Mrdalj, J., & Pallesen, S. (2012). Forekomsten av seksuelle overgrep i et representativt befolkningsutvalg i Norge. *Tidsskr Norsk Psykologforening*, *49*, 951–957.

Stemple, L., & Meyer, I. H. (2014). The sexual victimization of men in America: New data challenge old assumptions. *American Journal of Public Health*, *104*(6), e19–e26.

Thoresen, S., & Hjemdal, O. (2014). *Vold og voldtekt i Norge: En nasjonal forekomststudie av vold i et livsløpsperspektiv* [Violence and rape in Norway: A national prevalence

study with a life course perspective.] Norwegian only. Oslo: Nasjonalt kunnskapssenter om vold og traumatisk stress.

Weare, S. (2018). 'Oh you're a guy, how could you be raped by a woman, that makes no sense': Towards a case for legally recognising and labelling 'forced-to-penetrate' cases as rape. *International Journal of Law in Context*, *14*(1), 110–131.

Weiss, K. G. (2010). Male sexual victimization: Examining men's experiences of rape and sexual assault. *Men and Masculinities*, *12*(3), 275–298.