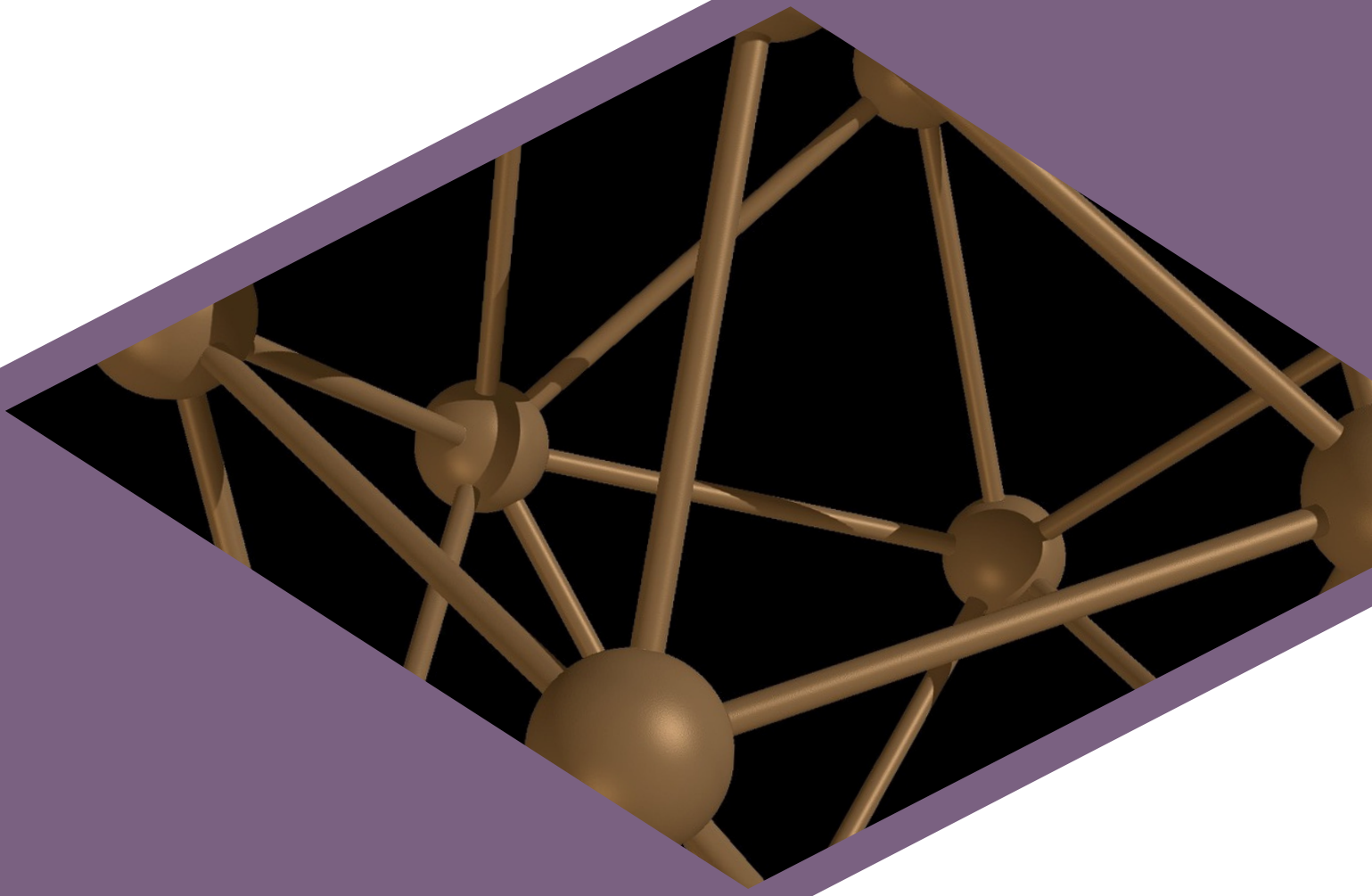
The Evidence Base



fsg

Trial methodologies

Clinical Trial Protocol

For reprint orders, please contact: reprints@futuremedicine.com

Journal of Comparative Effectiveness Research

Implementation of the *Connect for Health* pediatric weight management program: study protocol and baseline characteristics

Meg Simione*[1,2], Haley Farrar-Muir[1], Fernanda Neri Mini[1], Meghan E Perkins[1], Man Luo[1], Holly Frost[3,4], E John Orav[5], Joshua Metlay[6], Adrian H Zai[7], Caroline J Kistin[8,9], Kerry Sease[10,11], Simon J Hambidge[3,4] & Elsie M Taveras[1,2,12]; the Connect for Health Collaborative

[1]Division of General Pediatrics, MassGeneral Hospital for Children, Boston, MA 02114, USA
[2]Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA
[3]Primary Care – Pediatric Service, Denver Health, Denver, CO 80204, USA
[4]Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO 80045, USA
[5]Department of Medicine, Brigham & Women's Hospital, Boston, MA 02114, USA
[6]Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
[7]Laboratory of Computer Science, Massachusetts General Hospital, MA 02114, USA
[8]Department of Pediatrics, Boston University School of Medicine, Boston, MA 02118, USA
[9]Department of Pediatrics, Prisma Health, Greenville, SC 29605, USA
[10]Prisma Health, Greenville, SC 29605, USA
[11]Department of Pediatrics, University of South Carolina School of Medicine, Greenville, SC 29605, USA
[12]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
*Author for correspondence: Tel.: +1 617 643 0731; msimione@mgh.harvard.edu

We are implementing *Connect for Health*, a primary care-based intervention to improve family-centered outcomes for children, ages 2–12 years, in organizations that care for low-income children. We will use the 'Reach-Effectiveness-Adoption-Implementation-Maintenance' framework to guide our mixed-methods evaluation to examine the effectiveness of stakeholder-informed strategies in supporting program adoption and child outcomes. We also describe characteristics of children, ages 2–12 years with a BMI ≥85th percentile and obesity-related care practices. During the period prior to implementation, 26,161 children with a BMI ≥85th percentile were seen for a primary care visit and a majority lacked recommended diagnosis codes, referrals and laboratory evaluations. The findings suggest the need to augment current approaches to increase uptake of proven-effective weight management programs. **Clinical trial registration number:** NCT04042493 (Clinicaltrials.gov), Registered on 2 August 2019; https://clinicaltrials.gov/ct2/show/NCT04042493.

First draft submitted: 18 March 2021; Accepted for publication: 7 May 2021; Published online: 24 May 2021

Keywords: childhood obesity • electronic health records • health equity • implementation strategies • primary care

Background & rationale

Childhood overweight and obesity place a substantial burden on morbidity and quality of life and represent a national health priority [1–3]. The prevalence of childhood overweight and obesity remain at historically high levels and socioeconomic disparities appear to be widening [4–7]. Many of the underlying causes of obesity are modifiable risk factors throughout the life course; these risk factors represent major causes of health inequalities [8]. Approaches for reduction of obesity include collaborative interventions that aim to engage and empower families in obesity management and work across primary care and community settings [9]; however, adoption of interventions in these settings are limited.

The primary care setting provides an opportunity to detect elevated BMI levels and provide interventions that can alter a child's risk for disease and poor health outcomes. The US Preventive Services Task Force (USPSTF) guidelines offer strong evidence for screening and evaluation, counseling for weight management, a balanced

Future Medicine

nutrition plan and physical activity, and behavioral management techniques for lifestyle changes [10,11]. Yet, the USPSTF recommendations are not routinely followed and children with obesity are seldom identified [12]. It is critical that programs address the socio-contextual factors that affect behaviors at multiple levels including the individual, family and environment to improve health outcomes [13–15].

The *Connect for Health* pediatric weight management program is a novel approach to care delivery that leverages clinical and community resources to improve family-centered outcomes for high-risk children with overweight or obesity. The *Connect for Health* trial examined the comparative effectiveness of two clinical-community interventions in improving child BMI z-scores and family-centered outcomes and enrolled 721 children, ages 2–12 years with BMI ≥85th percentile in MA [16,17]. Children were randomized to one of two arms: enhanced primary care, e.g., flagging of children with BMI ≥85th percentile, clinical decision support tools, parent educational materials, neighborhood resource guide and text messages; or enhanced primary care plus contextually tailored, individual health coaching. At the end of the one-year intervention, both intervention arms resulted in improved family-centered outcomes and child BMI; there were no significant differences in outcomes between the two intervention arms [16].

The purpose of this study is to examine the implementation of the *Connect for Health* program across four organizations that deliver care to low-income children in the USA who have disproportionately high prevalence of obesity. We describe the study design, the mixed-methods evaluation plan and baseline characteristics and clinical care of children with obesity receiving care across the organizations. We present the study protocol in conjunction with the baseline characteristics to provide a comprehensive overview of the implementation settings, provide a roadmap for other organizations with similar characteristics and patient demographics, and stress the need for programs such as *Connect for Health*.

Design

Overview of study design

We used the Consolidated Framework for Implementation Research to assess contextual determinants in preparation of implementation of the *Connect for Health* pediatric weight management program in four organizations that deliver primary care to low-income children in Boston, MA, Denver, CO and Greenville, SC [18]. The *Connect for Health* program includes: electronic health record (EHR)-based clinical decision support tools to guide clinicians in weight management; family educational materials; text messages for parents to support behavior change. We have previously described the pre-implementation phase in which we engaged clinician and parent stakeholders to assess needs and preferences of the program tools and implementation strategies; and to identify barriers and facilitators to adoption [19]. Following stakeholder engagement, we iteratively adapted the program components to suit the implementation contexts, as well as in consideration of sustainability and scalability. We used the 'Reach-Effectiveness-Adoption-Implementation-Maintenance' (RE-AIM) framework to guide the mixed-methods evaluation of the program's implementation [20]. Using a quasi-experimental design, we will examine the effectiveness of stakeholder-informed strategies in supporting program adoption and child outcomes. At baseline (i.e., 15-months prior to program implementation), we abstracted EHR data from the four organizations to describe characteristics of children, ages 2–12 years with a BMI ≥85th percentile. Figure 1 illustrates the conceptual model for the implementation of the *Connect for Health* program, which guided our implementation strategies and evaluation plan. The study was registered at Clinicaltrials.gov (NCT02124460) and the Partners Healthcare institutional review board approved this study. The standard protocol items: recommendations for interventional trials for clinical trial study protocols and the standards for reporting implementation studies reporting guidelines (Supplementary Files 1 & 2) were followed.

Setting, participants, & end-users of the program

The *Connect for Health* program is being implemented in 26 primary care practices of four geographically and demographically diverse healthcare organizations. The organizations include: Boston Medical Center and Massachusetts General Hospital in Boston, MA, Denver Health in Denver, CO and Prisma Health in Greenville, SC. We selected the organizations because they have pediatric or family-medicine practices that are hospital-based, federally qualified or community health centers that deliver care to racially-ethnically diverse, low-income population of children with high rates of obesity. All the healthcare organizations use the Epic EHR platform (Verona, WI) allowing for the rapid scaling of EHR tools. Boston Medical Center is an academic medical center and is the largest safety-net hospital in New England. Massachusetts General Hospital is an academic medical center in Boston,

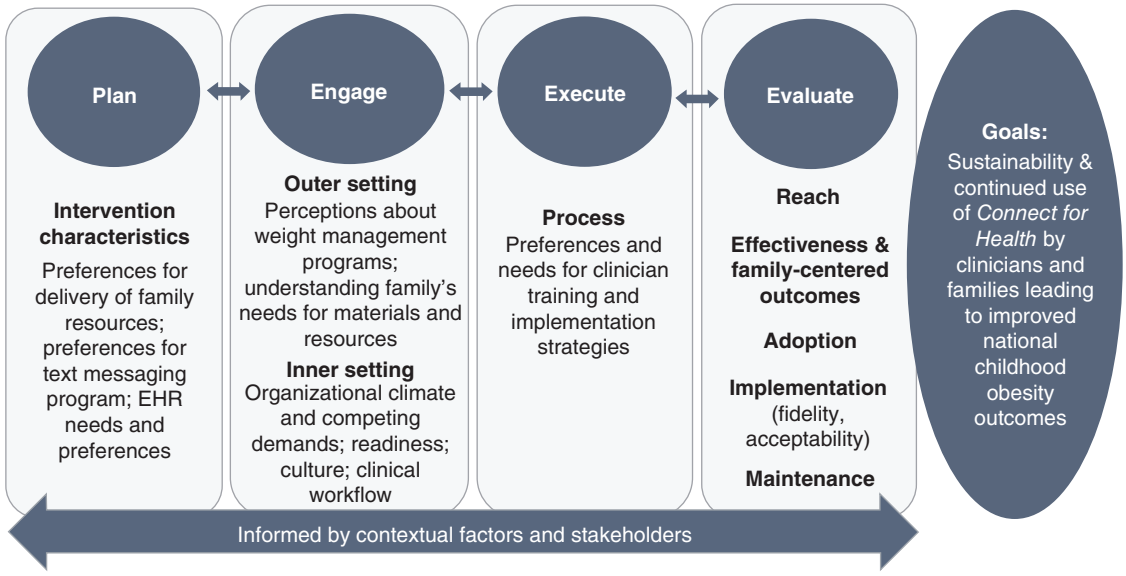


Figure 1. Implementation and Evaluation Approach for *Connect for Health* pediatric weight management program.
EHR: Electronic health record.

MA and has community health centers in surrounding cities. Denver Health is an academic health system, CO's primary safety-net institution, and the eighth largest federally qualified health center system in the US. Prisma Health is the largest multiregional health organization in SC.

The implementation of the *Connect for Health* program and its strategies are targeted toward pediatric or family-medicine primary care clinicians and is intended to be delivered during annual well-child visits or follow-up visits with the primary care team. Due to varying clinical workflows across the four healthcare organizations, physicians, physician assistants, nurse practitioners and medical assistants will use the program tools. Children, ages 2–12 years, with an elevated BMI and their families are the end-users of the program. During the preimplementation phase, each healthcare organization, based on their clinical population and needs, decided whether to make the program tools available for children with a BMI ≥85th or 95th percentile.

The Connect for Health program tools

Clinical-facing tools

The clinical decision support tools guide screening and management of childhood obesity. We created a Best Practice Alert (BPA), a flagging system that activates in the EHR for programmable patient specific characteristics that identify children with an elevated BMI at the time of a well-child visit. After a child's height and weight are taken and the data are entered into the EHR, a noninterruptive BPA appears to alert the clinician and/or staff to the elevated BMI. In addition to the BPA, we designed a SmartSet (an Epic visit template functionality) to assist clinicians in the best management practices for childhood obesity. The SmartSet prompts clinicians to document a diagnosis of overweight or obesity; discuss and document counseling on nutrition and physical activity; order laboratory evaluations as appropriate; make referrals to nutrition, weight management programs and other relevant services; place an order for the text-messaging program; provide educational materials and schedule a follow-up visit.

Family-facing tools

The family materials include a comprehensive set of printable patient educational handouts focusing on recommended behavioral changes that were adapted from the original trial based on stakeholder input [16,17,21,22]. The materials include an overview handout with the six behavioral messages and additional handouts focusing in-depth on each individual message. The messages include: healthy drink choices, screen-time, physical activity, following a balanced nutrition plans, sleep and social-emotional wellness. The tools also include an extensive library of social- and community-informed text messages to support behavior change. Clinicians and staff will enroll parents to receive the unidirectional, automated messages generally twice a week for 1 year. The community resource guides

Table 1. Characteristics of implementation strategies used to increase adoption of the *Connect for Health* program among pediatric primary care clinicians and staff.

Implementation strategy	Operationalizing the implementation strategies				Implementation outcome affected
	Actor	Action	Temporality	Dose	
1. Conduct ongoing training	Clinician champion; practice coach	Conduct trainings that focus on need for the program, evidence strength of the program and intervention components	Prior to program launch and ongoing throughout the implementation phase	Two trainings prior to program launch and then quarterly	Program uptake and fidelity
2. Provide local technical assistance and consultation	Clinician champion; practice coach; Epic analyst	Provide assistance in-person, over the phone and via email	Throughout the implementation phase	Ongoing as needed	Program uptake, feasibility and fidelity
3. Create a virtual learning community	Implementation support team	Provide education on the program and childhood obesity topics led by experts and offer continuing educational units	Will begin mid-way through the implementation period and last for 6–9 months	New module to be released monthly	Program uptake and fidelity
4. Alter incentive/allowance structures	Implementation support teams in conjunction with administrative leaders	Align program with healthcare organization's internal performance metrics and provide quality improvement bonuses	Throughout the implementation phase. Alignment with internal performance metrics that begins during the pre-implementation phase when adapting the program	Evaluated for qualification for bonus once during implementation phase	Program uptake, acceptability, and sustainability
5. Audit and provide feedback	Clinician champion; practice coach; implementation support team	Collect individual and practice-level metrics on utilization of the clinical decision support tools and deliver feedback reports to clinicians	Throughout the implementation phase	Feedback reports to be delivered quarterly	Program uptake
6. Facilitation	Clinician champion	Support and problem-solve with clinicians to encourage program adoption	Throughout the pre-implementation and implementation phase	As needed	Program uptake, acceptability and fidelity

assist families in identifying resources within their community that support behavior change. The community resource guides include sections on nutrition and food resources, physical activity and after-school programs, housing and utilities, and social services and healthcare. The family-facing materials have been translated into Spanish and Haitian Creole to ensure the program is accessible for the diverse communities that the four healthcare organizations serve. Besides materials being provided to families at their well-child visit, families can also obtain the patient educational materials and community resource guides from the *Connect for Health* website (www.c4hprogram.com).

Implementation strategies

The *Connect for Health* implementation strategies are designed to have an equity focus and support clinicians in the adoption of the program in primary care. During the implementation phase, each healthcare organization identified clinician champions, practice coaches and an implementation support team consisting of an Epic analyst and project manager. The implementation strategies are listed and operationalized according to the Expert Recommendations for Implementing Change [23,24] in Table 1. The strategies, such as conducting ongoing trainings and creating a virtual learning community, focus on educating clinicians about the program and best practices for screening and management of childhood obesity. Virtual learning communities have been widely used to increase knowledge and support practice change [25,26]. To support clinicians and staff, we will provide ongoing technical assistance to support their usage of the new EHR tools and other program components. Ongoing education and consultation are critical to provider adoption of clinical innovations and have been shown to be even more important than standalone training [23,27]. Clinician champions have been shown to facilitate change efforts by building organizational support [28] and by providing performance feedback that can support the adoption of evidence based practices [23,29–31]. To incentivize the uptake of the program, we aligned the program with each healthcare organization's internal performance metrics and when available, with quality improvement bonuses [32].

Outcome measures & evaluation

The RE-AIM framework has guided our evaluation and Table 2 shows our outcomes, measures and data sources.

Table 2. Study outcomes using the RE-AIM framework.

RE-AIM component	Measure	Data source
Reach	Child socio-demographic characteristics	EHR
	Rate of action taken on best practice alert among total number of best practice alerts fired	EHR
Effectiveness and family-centered outcomes	Change in BMI	EHR
	Family's experience with program	Parent survey administered within 8 weeks of well-child visit
Adoption	Setting-level characteristics (including number of practices, practice type)	Administrative data
	Staff-level characteristics (including clinicians and team members' role)	Administrative data
	Rate of Smart Set utilization and text messaging orders	EHR
Implementation		
Fidelity	Intervention & implementation fidelity checklist	Observation and interviews completed mid-implementation with clinicians, clinician champions and practice coaches
Acceptability	Acceptability of Intervention Measure	Survey administered mid-implementation to clinicians
Maintenance	Reach, effectiveness and adoption measures over time	EHR
	Clinical Sustainability Assessment Tool	Survey administered to unit chiefs, clinician champions and practice coaches

EHR: Electronic health record; RE-AIM: Reach-Effectiveness-Adoption-Implementation-Maintenance.

We will collect measures through EHR abstractions, surveys and informal interviews with leadership, clinician champions, practice coaches, clinicians and parents. To understand program reach, we will describe children's socio-demographic characteristics and will calculate the rate of action taken on the BPA among the total number of BPAs that were fired. We will measure adoption by describing setting- and staff-level characteristics and will report on SmartSet utilization and text messaging program orders. For implementation outcomes, we will assess fidelity to ensure the program is being delivered as intended with all core program components, and will measure program acceptability using the Acceptability of Intervention Measure [33]. We will evaluate reach, effectiveness and adoption measures over time to study maintenance, and will use the Clinical Sustainability Assessment Tool [34] to understand needs for program sustainment. We will calculate descriptive statistics for the reach, adoption, implementation and maintenance outcomes.

To understand effectiveness, we will examine changes in BMI z-score and family-centered outcomes over the course of program implementation. We will survey parents of eligible children eight weeks following their well-child visit to understand their experiences with the program. The survey, offered in English and Spanish, will include questions regarding how the program impacted behaviors and usefulness of the family-facing program tools. We will report on descriptive statistics of the survey.

We will use a quasi-experimental design to assess changes in BMI z-scores. We selected this design because we did not have enough sites for cluster randomization, randomizing within sites would have risked contamination, and the program has previously been shown to be effective and withholding the intervention would have not been ethical. Using only children who are eligible for the program, we will start with simple analyses that compare paired baseline and follow-up outcomes for each child. The baseline period will be 15 months prior to program implementation in which we will collect 2–3 measurements most proximal to the start of the program. We will also collect all BMI z-scores after the start of the program. A paired t-test will be used to compare the difference in the average pre-intervention and post-intervention BMI z-scores. While this approach assures the absence of confounders and excellent power, we have no control group. Any improvements that we observe could be present in other children, and therefore, not attributable to the intervention. Therefore, a regression discontinuity design will be used to assess program effectiveness by evaluating the reported changes in BMI z-scores. BMI z-scores will be collected pre-implementation and post-implementation for two groups of children: children who are eligible to receive the program (BMI $\geq$85th or 95th percentile) and 'quasi-control' children who are not eligible to receive the program (BMI between the 50th–85th percentile). From the observed pattern of changes in BMI z-scores in the quasi-control children, we can project what the BMI z-scores would be in children with an elevated BMI. We will then compare this projected pattern to the actual pattern in the children eligible to receive the program. A segmented regression model will be used to estimate the pattern in the control children, any acute change affecting

all children receiving the program equally, and any change in pattern that could affect the children with greater BMIs differentially.

Using an alternate design, we will evaluate changes in child BMI z-score by using a second control group of children with elevated BMIs at geographically and demographically matched community health centers. The inclusion of this secondary control group will allow us to match on BMI trajectories and will allow us to compare changes to BMI z-scores using a difference-in-differences design. We will collect data from community health centers through the Azara Healthcare Data Reporting and Visualization System (MA, USA). We will use multivariable linear regression models, adjusted for correlation due to repeated measures over time, to evaluate changes in BMI z-scores.

Baseline characteristics & obesity-related care metrics

To characterize our reach and target population, as well as understand current practices in obesity-related care at the healthcare organizations, we abstracted data from the EHR and collected the following information during the 15-month time period prior to program implementation: socio-demographics, BMI, BMI z-scores, BMI category (i.e., overweight, obesity and severe obesity), International Classification of Diseases, 10th Revision (ICD-10) codes for documentation of BMI, childhood obesity and nutrition and physical activity counseling, laboratory orders, referrals and the comorbid condition of asthma. The abstraction included children, ages 2–12 years with a BMI ≥85th percentile who were seen for a well-child visit at a practice implementing the program. The healthcare organizations implemented the program at different times, so the dates of the baseline periods differ. At primary care visits, we collected childhood obesity and nutrition and physical activity counseling ICD-10 codes. Laboratory orders included fasting glucose, hemoglobin A1c, ALT, AST and complete lipid panels. We searched laboratory orders completed at the time of well-child visits or during the visits 15-months prior. We included referrals to nutrition and weight management programs that were made at the time of well-child visits or during the visits 15-months prior. Referral data from Prisma Health were not available. We documented if a child had asthma as indicated by a prescription for albuterol, ICD-10 code, asthma referrals, historical registration of asthma or an asthma control test. For the laboratory orders, referrals and asthma documentation, when available, we also searched historical data as orders and referrals are not always recommended on a yearly basis. Historical data were limited due to availability in data warehouses. We calculated descriptive statistics for all the variables for each healthcare organization. Statistical analyses were completed using R Studio Software (version 3.5.1) and SAS (SAS Institute, NC, USA).

Results

During the 15-month period prior to implementation, 26,161 children with a BMI ≥85th percentile, ages 2–12 years were seen for a primary care visit. Estimated rates of childhood obesity across the organizations ranged from 35–50%. Across the four organizations, the mean (standard deviation [SD]) age of the children with a BMI ≥85th percentile was 7.8 (3.1) years and 49% of children were Hispanic, 22% were White and 18% were Black. Approximately 41% of families spoke a language other than English and 79% of children had public insurance. Table 3 shows the characteristics of children, ages 2–12 years with a BMI ≥85th percentile for the four healthcare organizations.

Table 4 shows the BMI, BMI z-score, BMI categories of children and obesity-related care metrics for children across the four healthcare organizations. Overall, approximately 48% of children had a BMI ≥95th percentile and 15% were in the severe obesity category defined as BMI ≥99th percentile. Between the organizations, the use of childhood obesity diagnosis codes and which family of codes was used (Z68 v. E66) varied, as well the usage between BMI categories. Most consistently, childhood obesity ICD-10 diagnosis codes were documented for children with severe obesity; the utilization for E66 codes was 60%. For children with obesity, the use was 44% and for children with overweight, the use was 17%. Counseling codes for nutrition and physical activity were not commonly used and usage was 7% for dietary counseling and 6% for physical activity. Orders placed for laboratory evaluations were more prevalent among children in higher BMI categories with the most orders being placed for the severe obesity category. For the overweight category, orders for all laboratory evaluations combined were 29%; for the obesity category 39% and for the severe obesity category 64%. Referrals placed for nutrition services and weight management programs increased between the BMI categories. Nutrition referrals were 4% for the overweight category, 7% for the obesity category and 16% for the severe obesity category. Weight management

Table 3. Characteristics of children, ages 2–12 years, with a BMI ≥85th percentile who were seen for a well-child visit during the 15-month period prior to program implementation (n = 26,161).

Child characteristics	Overall	Massachusetts General Hospital (September 2018–December 2019)	Boston Medical Center (July 2018–October 2019)	Denver Health (September 2018–December 2019)	Prisma Health (August 2018–November 2019)
	n = 26,161	n = 6752	n = 2494	n = 10,079	n = 6836
	n (%)	n (%)	n (%)	n (%)	n (%)
Age, mean (SD)	7.81 (3.14)	7.57 (3.25)	7.51 (3.27)	8.12 (3.04)	7.62 (3.08)
Sex					
Male	13,873 (53.03)	3583 (53.07)	1241 (49.76)	5453 (54.10)	3596 (52.60)
Female	12,288 (46.97)	3169 (46.93)	1253 (50.24)	4626 (45.90)	3240 (47.40)
Race/ethnicity					
Hispanic/Latino	12,923 (49.40)	3018 (44.70)	398 (15.96)	7579 (75.20)	1928 (28.20)
Non-Hispanic White	5786 (22.12)	1609 (23.83)	125 (5.01)	766 (7.60)	3286 (48.07)
Non-Hispanic Black	4585 (17.53)	664 (9.83)	1528 (61.27)	1206 (11.97)	1187 (17.36)
Non-Hispanic Asian	666 (2.55)	293 (4.34)	55 (2.21)	275 (2.73)	43 (0.63)
Non-Hispanic Other	675 (2.58)	479 (7.09)	40 (1.60)	136 (1.35)	20 (0.29)
Unknown	1526 (5.83)	689 (10.20)	348 (13.95)	117 (1.16)	372 (5.44)
Language	(n = 19,260)	(n = 6688)	(n = 2493)	(n = 10,079)	
English	11,337 (58.86)	4229 (63.23)	1629 (65.34)	5479 (54.36)	Not available
Spanish	6275 (32.58)	1998 (29.87)	152 (6.10)	4125 (40.93)	Not available
Other	1648 (8.56)	461 (6.89)	712 (28.56)	475 (4.71)	Not available
Insurance	(n = 20,085)	(n = 6731)	(n = 2451)		(n = 824)
Public insurance	15,945 (79.39)	4180 (62.10)	1961 (80.01)	9099 (90.28)	705 (85.56)
Private insurance	4140 (20.61)	2551 (37.90)	490 (19.99)	980 (9.72)	119 (14.44)

program referrals were 8% for the overweight category, 10% for the obesity category and 18% for the severe obesity category. Documentation of asthma ranged between 25 and 34% for the three BMI categories.

Conclusion

Pediatric primary care and community settings provide important opportunities to detect elevated BMIs, collaborate with families and deliver childhood obesity interventions. The *Connect for Health* pediatric weight management program is a scalable, proven-effective program that improves BMI and family-centered outcomes for children, ages 2–12 years. The program is being implemented in pediatric primary care practices of four healthcare organizations across the USA in which the majority of children are racially-ethnically diverse and low-income. We have described the study protocol for equity-focused implementation and evaluation and have described characteristics of children and obesity-related care metrics. During the 15-months prior to implementation, we found a low prevalence of guideline-adherent practices, including documentation of obesity and counseling codes, orders for laboratory evaluations and referrals for nutrition and weight management programs. The low uptake of these practices reinforces the importance of programs like *Connect for Health* being implemented in primary care.

The *Connect for Health* program was developed to follow the USPSTF guidelines and leverage clinical and community resources outcomes for children who are racially-ethnically diverse and from low-income communities given the persistent disparities in childhood obesity [5,11,16,17]. The results of the obesity-related care metrics demonstrated the opportunity to improve screening and interventions in the pediatric primary care setting. The USPSTF recommends screening for childhood obesity by calculating age- and sex-specific BMI [11]. Consistent with the literature, the use of childhood obesity diagnosis codes and exercise and counseling codes was low across the organizations resulting in missed opportunities to screen and document growth [35]. The documentation can also be reported to the Healthcare Effectiveness Data and Information Set allowing for accurate estimates of childhood obesity prevalence and trends. In their algorithm for childhood obesity assessment and management, the American Academy of Pediatrics recommends education, referrals to other healthcare providers and weight management programs and laboratory evaluations [36]. Despite these recommendations, uptake of screening, referrals and laboratory evaluations remain low as evidenced in our findings. Consistent with their algorithm, we found laboratory evaluations were ordered more often for children with obesity or severe obesity, as laboratory

Table 4. BMI and obesity-related care of children, ages 2–12 years, with a BMI ≥85th percentile who were seen for a well-child visit during the 15-month period prior to program implementation (n = 26,161).

	Overall	Massachusetts General Hospital (September 2018–December 2019)	Boston Medical Center (July 2018–October 2019)	Denver Health (September 2018–December 2019)	Prisma Health (August 2018–November 2019)
	n = 26,161	n = 6752	n = 2494	n = 10,079	n = 6836
	n (%)	n (%)	n (%)	n (%)	n (%)
BMI					
Mean (SD)	22.01 (4.43)	21.86 (4.27)	22.01 (4.60)	22.21 (4.38)	21.87 (4.57)
Z-score	1.77 (0.55)	1.76 (0.54)	1.78 (0.54)	1.76 (0.53)	1.78 (0.58)
BMI category					
– Overweight	12,484 (47.72)	3233 (47.88)	1163 (46.63)	4814 (47.76)	3274 (47.89)
– Obesity	9745 (37.25)	2566 (38.00)	952 (38.17)	3761 (37.32)	2466 (36.07)
– Severe obesity	3932 (15.03)	953 (14.11)	379 (15.20)	1504 (14.92)	1096 (16.03)
Childhood obesity diagnosis codes					
Overweight category					
– BMI 85th–95th percentile (Z68.53)	3030 (24.27)	190 (5.88)	5 (0.43)	1204 (25.01)	1631 (49.82)
– Diagnosis of overweight (E66.3)	2066 (16.55)	725 (22.42)	228 (19.60)	716 (14.87)	397 (12.13)
Obesity category					
– BMI ≥95th percentile (Z68.54)	4841 (49.68)	883 (34.41)	7 (0.74)	2183 (58.04)	1768 (71.70)
– Diagnosis of obesity (E66 codes)	4251 (43.62)	1415 (55.14)	655 (68.80)	1342 (35.68)	839 (34.02)
Severe obesity category					
– BMI ≥95th percentile (Z68.54)	2762 (70.24)	562 (58.97)	2 (0.53)	1219 (81.05)	979 (89.32)
– Diagnosis of obesity (E66 codes)	2363 (60.10)	782 (82.06)	339 (89.45)	587 (39.03)	655 (59.76)
Childhood obesity counseling codes					
Dietary counseling surveillance (Z71.3)	1824 (6.97)	155 (2.30)	4 (0.16)	207 (2.05)	1458 (21.33)
Exercise counseling (Z71.82)	1609 (6.15)	85 (1.26)	0 (0.00)	123 (1.22)	1401 (20.49)
Laboratory orders[†]					
Overweight category					
– Fasting glucose[‡]	865 (6.93)	0 (0.00)	0 (0.00)	1 (0.02)	864 (26.39)
– HgbA1c	1859 (14.89)	161 (4.98)	61 (5.25)	492 (10.22)	1145 (34.97)
– ALT	1934 (15.49)	709 (21.93)	17 (1.46)	598 (12.42)	610 (18.63)
– AST	662 (8.63)	33 (1.02)	18 (1.55)	Not available	611 (18.66)
– Lipid panel	1482 (11.87)	311 (9.62)	134 (11.52)	591 (12.28)	446 (13.62)
– Any lab order	3554 (28.47)	846 (26.17)	138 (11.87)	750 (15.58)	1820 (55.59)
Obesity category					
– Fasting glucose	894 (9.17)	2 (0.08)	6 (0.63)	4 (0.11)	882 (35.77)
– HgbA1c	2757 (28.29)	414 (16.13)	163 (17.12)	1106 (29.41)	1074 (43.55)
– ALT	2613 (26.81)	715 (27.86)	60 (6.30)	1168 (31.06)	670 (27.17)
– AST	830 (13.87)	101 (3.94)	58 (6.09)	Not available	671 (27.21)
– Lipid panel	2314 (23.75)	494 (19.25)	169 (17.75)	1106 (29.41)	545 (22.10)
– Any lab order	3755 (38.53)	834 (32.50)	194 (20.38)	1226 (32.60)	1501 (60.87)
Severe obesity category					
– Fasting glucose	561 (14.27)	2 (0.21)	9 (2.37)	7 (0.47)	543 (49.54)
– HgbA1c	2177 (55.37)	420 (44.07)	165 (43.54)	869 (57.78)	723 (65.97)
– ALT	1936 (49.24)	474 (49.74)	39 (10.29)	929 (61.77)	494 (45.07)
– AST	626 (25.78)	83 (8.71)	49 (12.93)	Not available	494 (45.07)
– Lipid panel	1953 (49.67)	440 (46.17)	163 (43.01)	883 (58.71)	467 (42.61)
– Any lab order	2505 (63.71)	552 (57.92)	185 (48.81)	954 (63.43)	814 (74.27)
Referrals[§]					
Overweight category	(n = 9210)				
– Weight management program	738 (8.01)	112 (3.46)	23 (1.98)	603 (12.53)	Not available
– Nutrition	395 (4.29)	298 (9.22)	16 (1.38)	81 (1.68)	Not available
Obesity category	(n = 7279)				
– Weight management program	747 (10.26)	147 (5.73)	105 (11.03)	495 (13.16)	Not available
– Nutrition	511 (7.02)	394 (15.35)	51 (5.36)	66 (1.75)	Not available

[†]Includes laboratory order placed at the time of well-child visit or prior to that visit. Historical data were available for MGH from June 2007 to December 2019, BMC from June 2018 to October 2019, Denver Health from January 2014 to December 2019 and Prisma Health from January 2011 to November 2019.
[‡]For Prisma Health, laboratory orders for serum glucose are shown.
[§]Includes referral order placed at the time of well-child visit or prior to that visit. Historical data were available for MGH from June 2015 to December 2019, BMC from June 2018 to October 2019 and Denver Health from April 2016 to January 2020.
[¶]Includes asthma documented at the time of well-child visit or prior to that visit. Documentation of asthma includes combination of albuterol prescriptions, ICD10 codes, asthma control test and historical registration of asthma. Historical data were available for MGH from August 2010 to December 2019, BMC from June 2018 to October 2019, Denver Health from December 2006 to December 2019 and Prisma Health from July 2009 to November 2019.
BMC: Boston Medical Center; HgbA1c: Hemoglobin A1c; MGH: Massachusetts General Hospital.

Table 4. BMI and obesity-related care of children, ages 2–12 years, with a BMI ≥85th percentile who were seen for a well-child visit during the 15-month period prior to program implementation (n = 26,161) (cont.).

	Overall	Massachusetts General Hospital (September 2018–December 2019)	Boston Medical Center (July 2018–October 2019)	Denver Health (September 2018–December 2019)	Prisma Health (August 2018–November 2019)
	n = 26,161	n = 6752	n = 2494	n = 10,079	n = 6836
	n (%)	n (%)	n (%)	n (%)	n (%)
Severe obesity category	(n = 2836)				
– Weight management program	498 (17.56)	170 (17.8)	104 (27.44)	224 (14.89)	Not available
– Nutrition	443 (15.62)	375 (39.35)	35 (9.23)	33 (2.19)	Not available
Asthma documentation¶					
– Overweight category	3177 (25.45)	946 (29.26)	297 (25.53)	1023 (21.25)	911 (27.83)
– Obesity category	2608 (26.76)	772 (30.09)	259 (27.21)	884 (23.50)	693 (28.10)
– Severe obesity category	1343 (34.16)	382 (40.08)	127 (33.51)	468 (31.12)	366 (33.39)

†Includes laboratory order placed at the time of well-child visit or prior to that visit. Historical data were available for MGH from June 2007 to December 2019, BMC from June 2018 to October 2019, Denver Health from January 2014 to December 2019 and Prisma Health from January 2011 to November 2019.
‡For Prisma Health, laboratory orders for serum glucose are shown.
§Includes referral order placed at the time of well-child visit or prior to that visit. Historical data were available for MGH from June 2015 to December 2019, BMC from June 2018 to October 2019 and Denver Health from April 2016 to January 2020.
¶Includes asthma documented at the time of well-child visit or prior to that visit. Documentation of asthma includes combination of albuterol prescriptions, ICD10 codes, asthma control test and historical registration of asthma. Historical data were available for MGH from August 2010 to December 2019, BMC from June 2018 to October 2019, Denver Health from December 2006 to December 2019 and Prisma Health from July 2009 to November 2019.
BMC: Boston Medical Center; HgbA1c: Hemoglobin A1c; MGH: Massachusetts General Hospital.

evaluations should only be ordered for children with overweight if they have risk factors present, and referrals should only be recommended after counseling with the primary care clinician. The *Connect for Health* program provides clinical decision support tools and clinician education, to screen, guide management practices and provide counseling in accordance with national guidelines to improve uptake of evidence-based practices.

The objective of this study is to increase adoption of the *Connect for Health* pediatric weight management program and evaluate the effectiveness of our implementation strategies. The study protocol we have presented is subject to potential challenges and limitations. As we implement the program, we will closely monitor program uptake and will take a practical approach by modifying our strategies and adapting as necessary. Throughout the implementation phase, we will document modifications to the program and implementation strategies using the framework for Reporting Adaptations and Modifications-Expanded [37,38]. We anticipate modifications to program delivery as three of the four healthcare organizations have shifted to telemedicine for well-child visits due to the COVID-19 pandemic. We selected pragmatic measurements for our evaluation plan; thereby we limited respondent surveys and selected outcomes that we could access through the EHR. Similarly, our baseline data pull was limited to variables within the EHR, as well as the availability of historical data (due to EHR vendor transitions) when searching for previously ordered laboratory evaluations, referrals and asthma documentation. For the current data abstraction, variables, including language, insurance and referral information were not consistently available resulting in missing data.

In conclusion, uptake of evidence-based practices for pediatric weight management fell well below expert recommendations across four organizations that deliver primary care to low-income children suggesting a substantial need for improving the delivery of high-quality care for children with obesity. Our findings emphasize the need to accelerate the adoption of proven-effective weight management programs particularly for children who are racially-ethnically diverse and from low-income households. The implementation of programs, such as *Connect for Health*, need to incorporate implementation strategies that address and advance child health equity.

Executive summary

- *Connect for Health* is a primary care-based intervention to improve family-centered outcomes for children, ages 2–12 years, in organizations that care for low-income children.
- The purpose of this study is to examine the implementation of the *Connect for Health* program across four organizations that deliver care to low-income children in the USA who have disproportionately high prevalence of obesity.
- This paper describes the study design, the mixed-methods evaluation plan and baseline characteristics and clinical care of children with obesity receiving care across the organizations, presenting the study protocol in conjunction

with the baseline characteristics to provide a comprehensive overview of the implementation settings, provide a roadmap for other organizations with similar characteristics and patient demographics, and stress the need for programs such as *Connect for Health*.

Author contributions

M Simione drafted the manuscript, conceptualized and designed the study, analyzed and interpreted the data, and drafted the initial manuscript. H Farrar-Muir, M Luo and ME Perkins analyzed and interpreted the data and critically reviewed the manuscript for important intellectual content. FN Mini, H Frost, EJ Orav, J Metlay, AH Zai, CJ Kistin, K Sease and SJ Hambidge assisted with interpretation of the data and critically reviewed the manuscript for important intellectual content. EM Taveras conceptualized and designed the study, interpreted the data and critically reviewed the manuscript for important intellectual content. The C4H Collaborative assisted with data collection and program implementation. All authors read and approved the final manuscript.

Acknowledgments

The following individuals and institutions constitute the *Connect for Health* Collaborative. The contact email for the collaborative is connectforhealth@partners.org. Massachusetts General Hospital: EM Taveras, M Simione, AA Boudreau, L Fiechtner, J Metlay, AH Zai, C Torres, H Farrar-Muir, S Price, M Luo and FN Mini. Boston Medical Center: CJ Kistin, WG Adams and JG Fiore. Denver Health: SJ Hambidge, H Frost, J Moreland, J Wallace, D Harr, S Sheck, S Raghunath, R Conrad and M Merithew. Prisma Health: K Sease, J Cassidy and C Craddock.

Financial & competing interests disclosure

Research reported in this article was partially funded through a Patient-Centered Outcomes Research Institute (PCORI) Award (DI-2017C3-9005) that is funding a clinical trial at Denver Health, Massachusetts General Hospital and Prisma Health; and through support by grant no. 1R01HL146782-01 from the National Heart, Lung and Blood Institute awarded to EM Taveras. EM Taveras time was supported by grant no. 1K24DK105989 from the National Institute of Diabetes and Digestive and Kidney Diseases. The statements in this article are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee and the National Institutes of Health. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The study protocol was approved by the Partners Health Care institutional review board.

Data sharing statement

The datasets used during the current study are available from the corresponding author on reasonable request.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Dietz WH. Overweight and precursors of Type 2 diabetes mellitus in children and adolescents. *J. Pediatr.* 138(4), 453–454 (2001).
2. Rankin J, Matthews L, Cobley S *et al.* Psychological consequences of childhood obesity: psychiatric comorbidity and prevention. *Adolesc. Health. Med. Ther.* 7, 125–146 (2016).
3. Kelsey MM, Zaepfel A, Bjornstad P, Nadeau KJ. Age-related consequences of childhood obesity. *Gerontology* 60(3), 222–228 (2014).
4. Hales CM, Fryar CD, Carroll MD, Freedman DS, Ogden CL. Trends in obesity and severe obesity prevalence in US youth and adults by sex and age, 2007–2008 to 2015–2016. *JAMA* 319(16), 1723–1725 (2018).

• **Provides trends in obesity prevalence, which underscores that childhood obesity is a public health crisis.**

5. Rossen LM, Schoendorf KC. Measuring health disparities: trends in racial-ethnic and socioeconomic disparities in obesity among 2- to 18-year old youth in the United States, 2001–2010. *Ann. Epidemiol.* 22(10), 698–704 (2012).
6. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity among adults and youth: United States, 2015–2016. *NCHS Data Brief.* (288), 1–8 (2017).

7. Olds T, Maher C, Zumin S *et al.* Evidence that the prevalence of childhood overweight is plateauing: data from nine countries. *Int. J. Pediatr. Obes.* 6(5–6), 342–360 (2011).
8. Marmot M, Friel S, Bell R, Houweling TAJ, Taylor S. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet* 372(9650), 1661–1669 (2008).
9. Dietz WH, Solomon LS, Pronk N *et al.* An integrated iramework for the prevention and treatment of obesity and its related chronic diseases. *Health Aff.* 34(9), 1456–1463 (2015).
10. O'Connor EA, Evans CV, Burda BU, Walsh ES, Eder M, Lozano P. Screening for obesity and intervention for weight management in children and adolescents: evidence report and systematic review for the US Preventive Services Task Force. *JAMA* 317(23), 2427–2444 (2017).
11. Grossman DC, Bibbins-Domingo K, Curry SJ *et al.* Screening for obesity in children and adolescents: US Preventive Services Task Force Recommendation Statement. *JAMA* 317(23), 2417–2426 (2017).

•• **Provides updates recommendations for screening and management for childhood obesity.**

12. Dietz WH, Baur LA, Hall K *et al.* Management of obesity: improvement of health-care training and systems for prevention and care. *Lancet* 385(9986), 2521–2533 (2015).

•• **Discusses the lack of evidence-based pediatric weight management programs and provides a compelling rationale for why interventions for pediatric weight management are needed.**

13. Parise I. The built environment and obesity: you are where you live. *Aust. J. Gen. Pract.* 49(4), 226–230 (2020).
14. Stokols D, Allen J, Bellingham RL. The social ecology of health promotion: implications for research and practice. *Am. J. Health Promot.* 10(4), 247–251 (1996).
15. Breslow L. Social ecological strategies for promoting healthy lifestyles. *Am. J. Health Promot.* 10(4), 253–257 (1996).
16. Taveras EM, Marshall R, Sharifi M *et al.* Comparative effectiveness of clinical-community childhood obesity interventions: a randomized clinical trial. *JAMA Pediatr.* 171(8), e171325 (2017).

•• **Provides the results of the primary outcomes of the Connect for Health comparative effectiveness randomized controlled trial.**

17. Taveras EM, Marshall R, Sharifi M *et al.* Connect for Health: design of a clinical-community childhood obesity intervention testing best practices of positive outliers. *Contemp. Clin. Trials* 45(Pt B), 287–295 (2015).
18. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement. Sci.* 4, 50 (2009).
19. Simione M, Frost HM, Cournoyer R *et al.* Engaging stakeholders in the adaptation of the Connect for Health pediatric weight management program for national implementation. *Implement. Sci. Commun.* 1(1), 55 (2020). https://doi.org/10.1186/s43058-020-00047-z

• **Provides the results of the stakeholder engagement for the implementation of Connect for Health and adaptations made to the program.**

20. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am. J. Public Health* 89(9), 1322–1327 (1999).
21. Sharifi M, Dryden EM, Horan CM *et al.* Leveraging text messaging and mobile technology to support pediatric obesity-related behavior change: a qualitative study using parent focus groups and interviews. *J. Med. Internet Res.* 15(12), e272 (2013).
22. Bala N, Price SN, Horan CM, Gerber MW, Taveras EM. Use of telehealth to enhance care in a family-centered childhood obesity intervention. *Clin. Pediatr. (Phila).* 58(7), 789–797 (2019).
23. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. *Implement. Sci.* 8, 139 (2013).
24. Powell BJ, Waltz TJ, Chinman MJ *et al.* A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement. Sci.* 10, 21 (2015).
25. Nix M, McNamara P, Genevro J *et al.* Learning collaboratives: insights and a new taxonomy from AHRQ's two decades of experience. *Health Aff.* 37(2), 205–212 (2018).
26. Wells S, Tamir O, Gray J, Naidoo D, Bekhit M, Goldmann D. Are quality improvement collaboratives effective? A systematic review. *BMJ Qual. Saf.* 27(3), 226–240 (2018).
27. Beidas RS, Edmunds JM, Marcus SC, Kendall PC. Training and consultation to promote implementation of an empirically supported treatment: a randomized trial. *Psychiatr. Serv.* 63(7), 660–665 (2012).
28. Bonawitz K, Wetmore M, Heisler M *et al.* Champions in context: which attributes matter for change efforts in healthcare? *Implement. Sci.* 15(1), 62 (2020).
29. Swindle T, Johnson SL, Whiteside-Mansell L, Curran GM. A mixed methods protocol for developing and testing implementation strategies for evidence-based obesity prevention in childcare: a cluster randomized hybrid Type III trial. *Implement. Sci.* 12(1), 90 (2017).
30. Ivers NM, Sales A, Colquhoun H *et al.* No more "business as usual" with audit and feedback interventions: towards an agenda for a reinvigorated intervention. *Implement. Sci.* 9, 14 (2014).

31. Ivers N, Jamtvedt G, Flottorp S *et al.* Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst. Rev.* (6), CD000259 (2012).
32. Murphy DJ, Lyu PF, Gregg SR *et al.* Using incentives to improve resource utilization: a quasi-experimental evaluation of an ICU quality improvement program. *Crit. Care Med.* 44(1), 162–170 (2016).
33. Weiner BJ, Lewis CC, Stanick C *et al.* Psychometric assessment of three newly developed implementation outcome measures. *Implement. Sci.* 12(1), 108 (2017).
34. CSAT. Clinical Sustainability Assessment Tool. https://sustaintool.org/csat/assess/
35. Taveras EM, Marshall R, Horan CM *et al.* Improving children's obesity-related health care quality: process outcomes of a cluster-randomized controlled trial. *Obesity (Silver Spring).* 22(1), 27–31 (2014).
36. American Academy of Pediatrics. American Academy of Pediatrics Institute for Healthy Childhood Weight. https://ihcw.aap.org/Pages/Resources_ClinicalSupports.aspx
37. Stirman SW, Baumann AA, Miller CJ, Wiltsey Stirman S, Baumann AA, Miller CJ. The FRAME: an expanded framework for reporting adaptations and modifications to evidence-based interventions. *Implement. Sci.* 14(1), 58 (2019).
38. Miller CJ, Barnett ML, Baumann AA, Gutner CA, Wiltsey-Stirman S. The FRAME-IS: a framework for documenting modifications to implementation strategies in healthcare. *Implement. Sci.* 16(1), 36 (2021).

Clinical Trial Protocol

Journal of Comparative Effectiveness Research

Protocol for Digital Real-world Evidence trial for Adults with insomnia treated via Mobile (DREAM): an open-label trial of a prescription digital therapeutic for treating patients with chronic insomnia

Frances P Thorndike*[,1], Richard B Berry[2], Robert Gerwien[1], Stephen Braun[1] & Yuri A Maricich[1]

[1]Pear Therapeutics, Inc., Boston, MA 02109, USA

[2]University of Florida Health Sleep Disorders Center, FL 32606, USA

*Author for correspondence: Tel.: +1 434 242 1282; frances.thorndike@peartherapeutics.com

Background: Cognitive behavioral therapy for insomnia (CBT-I) is underused in healthcare settings and is challenging for people with insomnia to access because of uneven geographical distribution of behavioral sleep medicine providers. Prescription digital therapeutics can overcome these barriers. This study evaluates the effectiveness of a specific digital CBT-I therapeutic. **Materials & methods: D**igital **R**eal-world **E**vidence trial for **A**dults with insomnia treated via **M**obile (DREAM) is a 9-week, open-label, decentralized clinical trial to collect real-world evidence for a digital therapeutic (Somryst™) delivering CBT-I to patients with chronic insomnia. The primary objective is to examine the effectiveness of Somryst to reduce self-reported insomnia symptoms and severity in a real-world population (n = 350). **Conclusion:** This pragmatic study seeks to assess the potential benefits of treating insomnia with an asynchronous, mobile, tailored prescription digital therapeutic.

Clinical trial registration: NCT04325464 (ClinicalTrials.gov)

Lay abstract: Chronic insomnia is linked to a range of health problems, including heart disease, chronic pain, high blood pressure and depression. A behavioral treatment called cognitive behavioral therapy for insomnia (CBT-I) is considered the first choice for helping patients overcome insomnia and reduce their risks of insomnia-related problems. Although the benefits of CBT-I have been established, it can be difficult for patients to access trained CBT-I therapists. One possible solution is to use digital forms of CBT-I, which patients can access on mobile devices. Somryst™ is a prescription digital therapeutic, which means it is authorized by the US FDA and has been proven effective in carefully-controlled clinical trials. Less is known, however, about how well the prescription digital therapeutic works in real-world settings. The Digital Real-world Evidence trial for Adults with insomnia treated via Mobile study (DREAM) will explore this question by evaluating a range of symptoms and outcomes in at least 350 patients with chronic insomnia who will use Somryst and be followed for 1 year.

First draft submitted: 5 January 2021; Accepted for publication: 10 February 2021; Published online: 8 March 2021

Keywords: CBT-I • chronic insomnia • cognitive behavioral therapy for insomnia • digital therapeutics • insomnia • PDT • prescription digital therapeutic

Background & rationale

Approximately 30 million Americans suffer from chronic insomnia [1], a condition associated with significant impairment in health and functioning [2–4]. Conditions co-occurring with insomnia include psychiatric disorders, heart disease, chronic pain and hypertension [5]. The annual medical and socioeconomic costs of chronic insomnia are estimated at US$30–70 billion [6].

Future Medicine

Cognitive behavioral therapy for insomnia (CBT-I) is the guideline-recommended, first-line treatment approach for adults with chronic insomnia [7–10], based on robust evidence of long-term efficacy and significantly lower safety risks compared with pharmacologic treatment options, which are associated with adverse events (AEs) and dependence [11,12]. CBT-I has been shown to improve insomnia itself as well as co-occurring symptoms [13], functional health, psychological well-being and sleep-related quality of life [14]. CBT-I, however, remains underused in healthcare settings and is challenging for people with insomnia to access [15,16]. Barriers to increased use of CBT-I include a paucity of healthcare providers trained in behavioral sleep medicine, inconvenience for patients due to the limited geographic diversity of CBT-I providers across the US, and patient concerns about privacy [16].

Prescription digital therapeutics (PDTs) represent a promising clinical approach for treating insomnia that can overcome the barriers currently constraining the use of CBT-I. A PDT for treatment of adults with chronic insomnia was market cleared in March 2020 by the US FDA [17,18]. The Somryst™ PDT is a native mobile application adaptation (with identical therapeutic content) of Sleep Healthy Using the Internet (SHUTi) – a web-based CBT-I intervention accessed via a responsive browser that has been extensively evaluated in at least nine randomized trials with over 3000 patients [19,20]. A meta-analysis of randomized trials found that the efficacy of digitally-delivered CBT-I is comparable with that delivered via traditional face-to-face modalities [21].

The objective of the Digital Real-world Evidence trial for Adults with insomnia treated via Mobile (DREAM) study is to examine the effectiveness of a digital therapeutic to reduce insomnia symptoms and severity in a real-world population of participants with chronic insomnia. The pragmatic study is designed to improve the understanding of how a PDT is actually used outside of a clinical trial setting and to assess its impact on insomnia symptoms and other measures of neuropsychiatric mental health and well-being.

Design

Study design

Prospective, open-label clinical trial with no comparator group.

Study setting

The DREAM study is decentralized. All participants will be recruited from within the US from a study waiting list that includes: participants who expressed interest in a mobile version of SHUTi when it became available; participants referred by their clinician; and participants who searched the internet for insomnia treatment. Participants may also be recruited using online advertising on social media sites.

Eligibility criteria

The DREAM study platform has been programmed with the inclusion and exclusion requirements allowing the system to automatically determine participant eligibility.

Inclusion criteria:

- Provision of electronic informed consent (prior to study-specific assessments).
- Age between 22 and 75 years, inclusively.
- Insomnia as defined by a score of 8 or above on the Insomnia Severity Index (ISI).
- Insomnia symptoms for at least 3 months [22].
- Access to a mobile device (i.e., smartphone or tablet) running supported versions of iOS or Android for the duration of the trial (including continuous data plan/Wi-Fi access).
- Resident of the US and living in the US for the duration of the trial.

Exclusion criteria:

- Presence of an active and progressive physical illness (e.g., congestive-heart failure, chronic obstructiNve pulmonary disease, or acute pain), neurological disorder (e.g., epilepsy) or neurological degenerative disease (e.g., dementia and multiple sclerosis).
- Unstable medication regimen (change to schedule or dosage within the past 3 months).
- Diagnosis of a psychotic disorder, bipolar disorder or a medical condition contraindicated by sleep restriction.
- Family or work schedules that interfere with normal sleep schedules (i.e., normal routine considered to be bedtime between 8:00 p.m. and 2:00 a.m. and/or waking times between 4:00 and 10:00 a.m.).

- Need to be alert or cautious to avoid serious accidents in a job or daily life. Examples include: long-haul truck drivers, long-distance bus drivers, air traffic controllers, operators of heavy machinery and some assembly line jobs.
- Pregnancy or intent to become pregnant during the trial.
- Other untreated sleep disorders as self-reported by the participant (e.g., obstructive sleep apnea, periodic leg movements, or parasomnias).
- Participation in an investigational research study in the past 30 days.

Eligible participants will be contacted via email and given an access code that will enable them to create an account and access the Somryst program. Ineligible participants will be notified via email and receive referral information to other sleep websites. In the event of uncertainty about a participant's eligibility, the Principal Investigator will contact the participant and/or participant's treating clinician to confirm if they are eligible for the study.

Who will take informed consent?

Participants who initially meet eligibility will complete an electronic informed consent form (eICF). Once the eICF is signed by both the participant and the Principal Investigator, the participant will be invited to complete additional medical/medication history questions, the ISI, and the Short Form 12 Health Survey to further determine eligibility and provide baseline data for the study. All data provided during the prescreening and screening periods is self-reported. The eICF will be reviewed and approved by the Institutional Review Board (IRB) prior to the start of the study. A copy of the IRB approval letter of the protocol, any amendments and the eICF will be filed in the Trial Master File.

Study procedures

Intervention description

The intervention to be evaluated in the study is an FDA-market-cleared PDT (Somryst) that delivers digital CBT-I therapeutic content to a patient via a mobile device (smartphone or tablet). CBT-I focuses on addressing the maladaptive behaviors, routines, and dysfunctional thoughts that perpetuate sleep problems, regardless of the cause of the sleep problems. Digital CBT-I is modeled on face-to-face CBT-I, which is typically delivered in weekly sessions over 6–8 weeks. The study program delivers 6 treatment Cores (learning modules), covering the following specific CBT-I therapy content:

1. **Get ready.** This Core sets the stage for the therapeutic experience. It lets the participant know what they will need to learn and do to improve their sleep and set goals for success.
2. **Sleep window.** This Core focuses on the concept of sleep restriction and consolidation. Using the participant's own data, the PDT will identify a tailored Sleep Window (a recommended Bedtime and Arising Time) that the participant should follow.
3. **Behaviors.** This Core focuses on stimulus control and works to break the connection between bed/bedtime and being awake.
4. **Thoughts.** This Core explains how a participant's thinking can contribute to chronic insomnia. The participant will learn to identify and shift these thought patterns to promote better sleep.
5. **Education.** This Core helps the participant identify changes to target in their lifestyle and environment to achieve better sleep.
6. **Look ahead.** This Core pulls together what the participant has learned, prepares participants for the future and teaches them what to do if they experience a relapse.

Cores must be completed sequentially and take approximately 30–45 min to complete. Each new Core is made available one week after the completion of the previous Core. Between Cores 1 and 2, at least 5 daily sleep diaries (integrated into the program) within a 7-day period must also be entered to unlock the next Core. Going forward, the participant must complete 5 out of 7 sleep diaries between Cores in order to receive an updated Sleep Window. Participants will have access to the program for 9 weeks, after which time their access will expire. Although all Cores can be completed in as little as 6 weeks, the intervention is made available for 9 weeks prior to postassessment to allow users sufficient time to access all Core materials, as well as implement new behaviors, strategies and techniques.

Table 1. Study objectives and end points.

Primary objectives	Primary end points
1. To examine the effectiveness of a digital therapeutic to reduce insomnia severity in a real-world insomnia patient population	1. Change in ISI from baseline to end of treatment
Secondary objectives	**Secondary end points**
1. Evaluate engagement and adherence rates with the digital therapeutic in a real-world patient population	1. Findings from in-therapeutic software application data: a. Core completion rate b. Intervention sleep diary completion rate c. Number of times the digital therapeutic is opened
2. Examine change in depression symptoms	2. Change in PHQ-8 score from baseline to end of treatment
3. Examine change in anxiety symptoms	3. Change in GAD-7 score from baseline to end of treatment
4. Examine change in insomnia severity, depression and anxiety at follow-ups	4. Change in ISI, PHQ-8 and GAD-7 scores from baseline to day 243 (~35 weeks) and day 428 (~61 weeks)
Exploratory objectives	**Exploratory end points**
1. Examine insomnia response to treatment	1. A decrease in ISI score from baseline to end of treatment >7 points
2. Examine insomnia remission	2. A final ISI score <8 points
3. Examine relationship among engagement and outcomes	3. ISI, PHQ-8 and GAD-7 findings from in-therapeutic software application data (e.g., Core completion or minutes spent in the PDT)
4. Examine relationship among assessments across time	4. ISI, PHQ-8 and GAD-7 results at specified time points
5. Evaluate participant satisfaction, usability, context and longitudinal experience and acceptance of the digital therapeutic and digital CBT-I	5. User Experience Surveys: NPS and SUS, qualitative diary data, participant interviews)
6. Determine if there is change in daytime sleepiness	6. Change in ESS from baseline to end of treatment
7. Determine change in quality of life	7. Change in SF-12 from baseline to end of treatment
8. Determine change in work attendance and productivity	8. Change in presenteeism/absenteeism work questions from baseline through end of treatment, day 243 and day 428

CBT-I: Cognitive behavioral therapy for insomnia; ESS: Epworth Sleepiness Scale; ISI: Insomnia Severity Index; NPS: Net Promoter Score; PDT: Prescription digital therapeutic; SF-12: Short Form 12 Health Survey; SUS: System Usability Scale.

Criteria for discontinuing or modifying allocated interventions

There are no special criteria for discontinuing or modifying allocated interventions. Participants may choose to stop using the Somryst therapeutic at any point and do not need permission to do so.

Strategies to improve adherence to interventions

The Somryst therapeutic automatically reminds users to complete treatment and study procedures via email and push notifications. Study team members will also send manual emails to remind participants of study procedures during the study.

Outcomes

Table 1 details the primary, secondary and exploratory outcomes of the DREAM study and the assessment tools used to evaluate each end point.

Participant timeline

Participants will complete assessments at end of treatment (9 weeks/day 63), 6 month follow-up (day 243) and 12 month follow-up (day 428) (see Table 2 for details). A subset of approximately 34 participants will be asked to partake in an optional user experience sub study (see Table 3). Participants will be asked to sign an additional eICF for the sub-study. The user experience sub-study includes the following data collection components:

- Qualitative Diary Data collected from the participant over the course of 5 consecutive days starting in Core 2 and Core 4.
- Interviews conducted with the participant via Zoom interview at end of Core 4 and start of Core 6.

Sample size

This is an open access study and is expected to be overpowered for the primary end point based on previous trials with a similar product. A trial of Somryst's predecessor, SHUTi, reported an effect size of 1.90 or higher [19], which corresponds to a sample size of 6 (assuming 95% power at an alpha of 0.05). In order to better characterize

Table 2. Schedule of events.

Study phase	Prescreening	Screening	Treatment (Core)						End of treatment (Day 63)	Follow-up (Day 243)	Follow-up (Day 428)	Ref.
			1	2	3	4	5	6				
Procedures												
Informed consent		X										
Inclusion/exclusion	X	X										
Demography	X											
Medical/medication History		X							X	X	X	
Sleep restriction window				X	X	X	X	X				
Sleep diary			X	X	X	X	X	X				
ISI		X	X	X	X	X	X	X	X	X	X	[23]
PHQ-8		X	X		X		X		X	X	X	[24]
GAD-7		X	X		X		X		X	X	X	[25]
Epworth Sleepiness Scale		X		X	X	X	X	X	X	X	X	[26]
User Experience SUS & NPS						X	X					
Health economics and outcomes questions (survey)		X							X	X	X	
SF-12		X							X	X	X	[27]
Adverse events collection (self report)			X	X	X	X	X	X	X	X	X	

ISI: Insomnia Severity Index; NPS: Net Promoter Score; SF-12: Short Form 12 Health Survey; SUS: System Usability Scale.

Table 3. Optional user experience sub-study schedule of events.

Study phase	Treatment (Core)						End of treatment (Day 63)
	1	2	3	4	5	6	
Procedures							
Informed consent	X						
User experience diary data		X			X		
User experience interview				X		X	

the secondary end points, we expect to enroll approximately 350 participants. For the qualitative user experience substudy, a sample size of 34 participants was selected as that sample size should be adequate to reach saturation in the qualitative interviews [28].

Recruitment

As noted earlier, all participants will be recruited from the US from a study waiting list that includes: participants who previously used SHUTi (the web-based precursor of Somryst) who have expressed interest in the mobile version; participants referred by their clinician; and participants who searched the internet for insomnia treatment.

Data management

Oversight of data management, including electronic data collection, storage and export, security, tracking, data analysis and quality assurance will be the responsibility of Pear. The Principal Investigator will also be responsible for ensuring that all study staff adhere to human participants/IRB guidelines related to data management. Data files will be backed up at regular intervals and will be accessible only by trained study staff members.

Data collection

All participant data are collected within the study website and digital therapeutic and stored in the sponsor's database. Eligibility determinations and medical release waivers will be collected in the source records and maintained electronically. The eligibility determination will be documented in the study dashboard by the Principal Investigator or designee. All AEs will be collected by the contract research organization (CRO) and entered into an electronic case report form. The source data will be retained electronically at the clinical site until notification is given by Pear for destruction.

Monitoring

The study will be monitored remotely by Pear or its representative (the Study Monitor) to ensure it is conducted and documented according to the protocol, International Conference on Harmonisation/Good Clinical Practice (ICH/GCP) and all applicable regulatory requirements. The Investigator will work closely with the Study Monitor and, as needed, provide them appropriate evidence that the conduct of the study is being done in accordance with applicable regulations and GCP guidelines.

Quality assurance & quality control

Although Pear or its designee will perform the quality assurance and quality control activities of this study, responsibility for the accuracy, completeness and reliability of the study data presented to Pear will lie with the participants and Investigator. Prior to the study initiation, Pear will explain the protocol and instructions for using the study product to the Investigator assigned by the CRO. In addition, a Pear Clinical Operations Manager will be available to explain applicable regulations and to answer any questions regarding the conduct of the study. At its discretion, Pear may conduct audits as part of the implementation of quality assurance to ensure that the study is being conducted in compliance with the protocol, standard operating procedures, GCP and all applicable regulatory requirements. The study center may also be compelled to an inspection by a Regulatory Authority.

Source data

Source data are defined as information in original records and certified copies of original records of clinical findings, observations, data or other activities in a clinical study necessary for the reconstruction and evaluation of the study. In this decentralized trial, data collected through the Somryst mobile program infrastructure and study website will be considered as source. In addition, source documentation will be maintained by the Principal Investigator for determination of participant eligibility and safety reporting.

Record keeping

The Investigator must arrange for retention of study records (Essential Documents for the Conduct of a Trial are listed in the ICH Guideline for Good Clinical Practice) at the site, in a secure location. Records will be kept for the duration of the product lifetime, per Pear's quality management system record retention requirements, or a period of 2 years after completion of the study, whichever is longer. These documents will be retained for a longer period, however, if required by the applicable regulatory requirements or by agreement with Pear. It is the responsibility of Pear to inform the Investigator/institution as to when these documents no longer need to be retained. The Investigator will take measures to prevent any accidental or premature destruction of these documents.

Confidentiality

The protection of patient privacy is of the utmost importance and is a principal consideration throughout the software design. The approaches to data security undertaken were based on a number of recommendations, principally National Institute of Standards and Technology Cybersecurity Practice Guide, Special Publication 1800–1: "Securing Electronic Health Records on Mobile Devices," and on guidance from a number of consultants and vendors.

Data transport & storage

All data collected by the device are hosted and stored in Amazon Web Services (AWS). AWS follows a variety of internationally recognized security standards such as the National Institute of Standards and Technology SP800–53 and Health Insurance Portability and Accountability. All participant information is automatically encrypted by AWS when it is entered into the system, allowing for secure data transfer and storage. All back-end services operate

in nonroutable IP space on dedicated instances, are only accessible to external callers through secured, software mediated interfaces and are only accessible via Hypertext Transfer Protocol Secure. All communications between subsystems (Data in Motion) are encrypted using Transport Layer Security. No access is provided to any subsystem via unencrypted protocols. All communication between the backend, service layer and the mobile application is further secured through Advanced Encryption Standard-encrypted session tokens, used to identify all actors in the system without providing any data visibility to any third party, intentionally or otherwise.

Data stored in the service layer are encrypted according to the Advanced Encryption Standard. A cryptographically secure pseudo-random number generator is used to generate the encryption key for mobile app storage. All use data within the mobile application are stored in an encrypted queue and flushed to the Backend Services securely over Hypertext Transfer Protocol Secure whenever internet connectivity is available.

Data segregation

In order to limit access to Protected Health Information (PHI) and Personally Identifiable Information (PII), we maintain completely isolated infrastructure for the Production Environment and provide a functionally identical Development, Staging and QA environment for software development, user story acceptance and testing. Access to the Production Environment is limited to quality controlled and audited software developed by the Pear Therapeutics team and is further limited to a set of authorized users, with limited access for environment management and maintenance purposes. Production Data, including all PHI and PII, is maintained in the Production Environment only; it cannot enter the Development, Staging and QA environment, nor can it be copied for any purpose to desktop or laptop computers used for development. These environments are maintained in separate accounts, to prevent access to anyone other than minimum authorized administrators, and to prevent system-level attacks or inadvertent access to PHI or PII.

Statistical methods

Statistical methods for primary & secondary outcomes

The Intention-to-Treat population will serve as the primary population for the analysis of efficacy and safety data in this trial. The Intention-to-Treat consists of all participants who have a baseline observation for the analysis end point. The per-protocol set will include all participants who complete the 6 CBT-I cores.

Descriptive statistics will be used to evaluate the enrolled population and the population at risk for each of the 6 Cores as well as at end of treatment (day 63) and follow-up (day 243 and day 428). Categorical data will be presented as frequencies and percentages of participants at risk at baseline. For continuous data, mean, standard deviation, median, first and third quartile, minimum and maximum will be presented.

Primary end point analysis

The primary efficacy end point in this study is the ISI score measured at baseline, day 63, day 243 and day 428. The ISI score will be analyzed using a mixed-effects model for repeated measure (MMRM) including the fixed categorical effect of visit with subject as a random effect. An unstructured correlation matrix will be used to model the within-subject errors. The primary hypothesis to be tested is that the study PDT reduces insomnia severity, as measured by change in ISI, from baseline to end of treatment (day 63) and to follow-up assessments (day 243 and day 428). Significance will be evaluated with a two-tailed p-value of 0.05. To control for multiplicity end of treatment will be evaluated first and if significant days 243 and 428 will be evaluated. In addition, estimated marginal means ± 95% CIs will be calculated for each assessment point and Cohen's d will be used as an estimate of effect size for the change from baseline to each assessment point.

Secondary end point analysis

Engagement with the study PDT will be estimated from the rate of core completion, the rate of sleep diary completion and the number of times the PDT is opened. These variables will be summarized using descriptive statistics. As an exploratory end point, minutes spent in the PDT will also be captured. Change in patient health questionnaire 8 (PHQ-8) and general anxiety disorder 7 (GAD-7), will also be analyzed with an MMRM analysis in a manner consistent with ISI. The significance test will be carried out on a visit effect using a one-sided alpha of 0.05/2 = 0.025 to control for multiplicity if the primary end point is significant with end of treatment being evaluated first in both cases. In addition, estimated marginal means ± 95% CIs will be calculated for each visit and Cohen's d will be used as an estimate of effect size for each time point.

Qualitative user experience analysis

The qualitative interviews with participants are open-ended interviews. With participant permission, interviews are recorded and transcribed verbatim (single coder). A trained user experience researcher then serves as the coder of the transcribed interview to categorize the data gathered, aiming to discover major themes that emerge from the data. *A priori* codes are used to denote challenges, motivations and other contexts in the data. These processes follow grounded theory which provides a common data analysis method for qualitative data across human computer interactions [29–31].

Interim analyses

Interim analyses, after all participants complete post-treatment, may be conducted upon IRB review and approval of interim analysis plans.

Methods for additional analyses (e.g., subgroup analyses)

For exploratory end point analyses, insomnia responders (decrease in ISI score from baseline to end of treatment >7) and remitters (final ISI score <8) will be counted and tabulated. Correlations between engagement and clinical outcomes will be evaluated using both Pearson's correlation coefficient and Spearman's rank correlation as follows. Change from baseline (follow-up – baseline) will be calculated for ISI, PHQ-8, GAD-7 and Epworth Sleepiness Scale at both the end of treatment and the end of follow-up. These will be correlated with core completion rates, sleep diary completion rate and the number of times the PDT is opened. In addition, correlations among the engagement variables and among the clinical outcomes will also be evaluated.

User experience surveys, qualitative diary data and participant interviews will be summarized using descriptive statistics. Change in Epworth Sleepiness Scale and Short Form 12 Health Survey will be analyzed with an MMRM analysis consistent with ISI, PHQ-8 and GAD-7. All four clinical outcomes can be expressed as clinical categories of severity. For example, an ISI score of 22–28 corresponds to severe insomnia, 15–21 moderate insomnia, 8–14 subthreshold insomnia and 0–7 no insomnia. Similar categories exist for the three other clinical metrics. Clinical improvement in outcomes will be summarized using shift tables from baseline categories to the end of treatment and the end of follow-up.

Methods in analysis to handle protocol nonadherence & any statistical methods to handle missing data

Participants with missing data will initially be evaluated assuming data are missing at random in the MMRM. Sensitivity to this assumption will be evaluated with multiple imputation, last observation carried forward imputation and evaluation of the per protocol core 6 completers which should reduce the amount of missing data and will be detailed in the statistical analysis plan prior to database lock.

Oversight & monitoring

Composition of the coordinating center & trial steering committee

Operational activities and processes are completed through rigorous management and trial oversight of a CRO under supervision of a site Principal Investigator. In addition, a cross-functional team for the Sponsor meets weekly or bi-weekly to review study status, including recruitment, participant support, study milestones and any safety concerns that may arise.

Composition of the data monitoring committee, its role & reporting structure

A data monitoring committee is not required for this study because Somryst is an FDA-cleared medical device. All AEs and serious AEs (SAEs) and tracked and reported as detailed below.

AE reporting & harms

AEs will be self-reported by participants throughout the study, beginning at the time the participant gives informed consent through the last follow-up period. Participants will be provided with telephone and email contacts for the investigator and/or study support to address any health or technical related questions. Participants are also given instructions within the PDT that it is not intended for emergencies and to contact 911 or go to the nearest emergency room for all emergent concerns. The Investigator or designee and research site staff are responsible for the documentation, classification, reporting and follow-up of events meeting the definition of an AE or SAE.

Risks & benefits

Somryst therapeutic content has demonstrated significant benefits to participants with chronic insomnia and chronic insomnia with depression [19,20]. More specifically, compared with comparison conditions, participants experienced:

- Significant improvement in insomnia severity (as measured by clinically validated and standardized instruments, including the ISI).
- Significant improvement in symptoms of depression (as measured by clinically validated and standardized instruments, including the patient health questionnaire 9 [PHQ-9]).
- Significant reductions in symptoms of anxiety (as measured by clinically validated and standardized instruments, GAD-7).

The following risks are associated with use of Somryst:

- Digital concerns: Some participants may feel uncomfortable providing data electronically and may have concerns about the confidentiality of their digital data. They may also have concerns about the legitimacy of a digital therapeutic.
  Security measures to protect participant data include the fact that the PDT requires a username and password for access. Users set up a 4-digit PIN and have the option to grant permission to use biometric features (Face ID or Touch ID on iPhone, Fingerprint on Android phones).
- Sleep restriction (and consolidation) within Somryst can cause sleepiness, especially in the early stages of use. Increased daytime sleepiness is normal and expected, but it is also temporary. Sleepiness that persists after a few weeks of treatment may indicate the presence of another sleep disorder or medical condition other than insomnia.

Definitions

An AE is any untoward medical occurrence in a participant or clinical investigation and may not necessarily have a causal relationship with the administered treatment. An AE can therefore be any unfavorable and unintended sign (including a clinically significant laboratory abnormality, for example), symptom or disease temporally associated with the use of a treatment whether or not related to the treatment itself. Pre-existing conditions, diseases or disorders are not considered AEs unless there is a change in intensity, frequency or quality.

An SAE is any untoward medical occurrence that:

- Results in death.
- Is life-threatening (at the time of the event).
- Requires inpatient hospitalization or prolongation of existing hospitalization.
- Results in persistent or significant disability/incapacity.
- Is a congenital anomaly/birth defect.

Important medical events that may not result in death, be life-threatening or require hospitalization may be considered an SAE when, based upon appropriate medical judgment, they may jeopardize the participant and may require medical or surgical intervention to prevent one of the outcomes listed in this definition. A suspected unexpected serious adverse reaction is an SAE that is not identified in nature, intensity or frequency in the risk information set out in the risks and benefit section of this protocol.

Classification of AE intensity

The Investigator or designee is responsible for making an assessment as to the seriousness, intensity, causality and outcome of an AE. The Investigator will determine causality as either related or unrelated to Somryst. For each recorded AE or SAE, the investigator or designee must assess intensity based on the criteria listed in Table 4 and follow the classification schemes detailed in Tables 5 & 6. If there is insufficient information to determine intensity, the AE must still be reported.

AEs will be followed up per the site's standard operating procedures requirements.

Table 4. Adverse event classifications.

Classification	Definition
Grade 1 (mild)	An event that is usually transient and may require only minimal treatment or therapeutic intervention. The event does not generally interfere with usual activities of daily living.
Grade 2 (moderate)	An event that is alleviated with additional specific therapeutic intervention. The event interferes with usual activities of daily living, causing discomfort but poses no significant or permanent risk of harm to the participant.
Grade 3 (severe)	An event that requires intensive therapeutic intervention. The event interrupts usual activities of daily living or significantly affects clinical status. The event poses a significant risk of harm to the participant and hospitalization may be required.

Table 5. Classification of adverse event causality.

Classification	Definition
Unrelated	The AE or SAE is judged to be clearly and incontrovertibly due only to extraneous causes (e.g., disease, environment) and does not meet the criteria for study product relationship listed under probable, possible or unlikely.
Unlikely	The AE or SAE is unlikely related to the study product when the AE or SAE: • Does not follow a reasonable temporal sequence from administration of the study product • May readily have been produced by the participant's clinical state, environmental or toxic factors, drugs or other modes of therapy administered to the participant • Does not follow a known pattern of response to the study product • Does not reappear or worsen when the study device is re-administered
Possible	The AE or SAE is possibly related to the study product when the connection to the study product appears unlikely but cannot be ruled out with certainty. This causal relationship is assigned when the AE or SAE: • Follows a reasonable temporal sequence from administration of the study product • May have been produced by the participant's clinical state, environmental or toxic factors, drugs or other modes of therapy administered to the participant • Follows a pattern of response to the suspected study product
Related	The AE or SAE is probably related to the study product when the connection to study product can be made with a high degree of certainty. This causal relationship is assigned when the AE or SAE: • Follows a reasonable temporal sequence from administration of the study product • Cannot be reasonably explained by the known characteristics of the participant's clinical state, environmental or toxic factors, drugs or other modes of therapy administered to the participant • Disappears or decreases upon cessation or reduction in product use (note that there are important exceptions when an AE or SAE does not disappear upon discontinuation of the study product, yet product relatedness clearly exists) • Follows a known pattern of response to the suspected study product • Reappears upon re-challenge

AE: Adverse event; SAE: Serious adverse event.

Table 6. Classification of adverse event outcomes.

Classification	Definition
Fatal	The participant died
Resolved	The AE or SAE has ended
Resolved with sequelae	The AE or SAE has ended but changes are noted from baseline
Unresolved	The AE has not ended. And AE outcome can only be categorized as unresolved if the AE is: • Ongoing at the end of the reporting period after the final follow-up visit, and the investigator deems that further follow-up is not medically required • Lost to follow-up after repeated unsuccessful attempts to contact the participant • Ongoing and referred to the participant's physician or a specialist

AE: Adverse event; SAE: Serious adverse event.

SAE reporting

The Investigator is required to contact Pear Therapeutics, Inc. (Pear) within 24 h of learning of any SAE.

Pear Therapeutics, Inc.:

Telephone: +1 833 ASK-PEAR (275 7327).

Email: pearconnect@peartherapeutics.com

If the SAE is fatal or life threatening, Pear must be informed immediately. For reporting of all SAEs, Investigator must scan/email all completed pages of the SAE report form within 24 h to the Medical Monitor. To ensure participant safety, every SAE, regardless of causality, occurring after the participant has provided informed consent and until the last study visit must be reported to Pear within 24 h of learning of its occurrence.

All follow-up information for the SAE including information on complications, progression of the initial SAE and recurrent episodes must be reported as follow-up to the original episode within 24 h of the investigator receiving the follow-up information. An SAE occurring at a different time interval or otherwise considered completely unrelated to a previously reported one must be reported separately as a new event. Note: SAEs, related or possibly related, to Somryst are subject to the SAE reporting requirements in this section.

Frequency & plans for auditing trial conduct

Teams from the Sponsor and research site meet weekly or bi-weekly to discuss trial recruitment, participant support any IRB-related requirements, study logistics and any safety concerns should they arise. In addition, documentation required by the IRB is developed and reviewed as required, including annual reports on study progress.

Plans for communicating important protocol amendments to relevant parties (e.g., trial participants, ethical committees)

Substantive changes in the protocol include changes that affect the safety of participants, and/or changes that alter the scope of the investigation or the scientific quality of the study. Such changes must be prepared as a protocol amendment and approved by the IRB prior to implementation. If a protocol amendment requires changes in the eICF, the revised eICF must also be approved by the IRB.

Conclusion

This pragmatic study is designed to mimic a real-world scenario to understand how Somryst is used by individuals with insomnia outside of a randomized clinical trial setting. The number of clinical assessments has been minimized in terms of frequency and type to reduce clinical trial burden. Clinical scales and surveys will be used to evaluate a range of efficacy, usability and health-related outcomes that will help further an understanding of the impact of insomnia and the potential benefits of treating insomnia with an asynchronous, contact-less prescription digital therapeutic.

Trial status

Study no.: PEAR-003-101, version 3.0, 12 May 2020. The DREAM study is currently recruiting and plans to recruit until the projected sample size is met.

Summary points

- Cognitive behavioral therapy for insomnia (CBT-I) is the guideline-recommended first-line treatment for patients with chronic insomnia but it is underused in healthcare settings and is challenging for patients to access.
- Prescription digital therapeutics may help expand access to cognitive behavioral therapy for insomnia and, not only improve insomnia symptoms, but also be connected to improvement in co-occurring physical and psychological illnesses.
- The prescription digital therapeutic Somryst™ has been proven effective in randomized controlled trials and is US FDA-approved for patients with chronic insomnia, but less is known about its effectiveness in real-world settings.
- Digital Real-world Evidence trial for Adults with insomnia treated via Mobile (DREAM) is a 9-week, open-label, decentralized clinical trial to collect evidence for Somryst in a real-world population of 350 patients with chronic insomnia.
- Outcomes to be evaluated include: insomnia symptoms; engagement with and adherence to the therapeutic; symptoms of depression and anxiety; and rates of insomnia response and remission.
- Outcomes will be assessed at the end of treatment (week 9) and again at 6 and 12-month follow-ups.

Author contributions

FP Thorndike and YA Maricich conceived the study concept and design for the Digital Real-world Evidence trial for Adults with insomnia treated via Mobile study. RB Berry is the site investigator and contributed to study design. R Gerwien is the biostatistician for the study and will perform the statistical analyses of study results. S Braun is a medical writer who wrote the first draft of this article. All named authors adhere to the authorship guidelines of *JCER*. All authors have agreed to publication.

Financial & competing interests disclosure

This study is funded entirely by Pear Therapeutics, Inc. The sponsor oversaw the study design; collection, management, analysis and interpretation of data; writing of the report; and the decision to submit the report for publication. FP Thorndike, R Gerwien,

S Braun and YA Maricich are employees of Pear Therapeutics, Inc., which develops the Somryst™ digital therapeutic discussed in this protocol. RB Berry is an employee of the University of Florida. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. The Investigator and sub-Investigators, as noted on the US FDA Form 1572, will provide Pear Therapeutics Inc. with sufficient accurate financial disclosure information to allow Pear to maintain complete and accurate certification or disclosure statements as required under 21 CFR Part 54. The Investigator shall promptly update this information if any relevant changes occur during the investigation and for 1 year following the completion of the study.

Editorial assistance on the writing of this paper was provided by N Enman and S Edington (employees of Pear Therapeutics, Inc.).

Ethical conduct of research

This study will be conducted in full accordance with all applicable Policies and Procedures and all applicable US federal and state laws and regulations including 45 Code of Federal Regulations (CFR) 46, and the Health Insurance Portability and Accountability (HIPAA) Privacy Rule. Any episode of noncompliance will be documented. The Investigator will perform the study in accordance with this protocol and will report unexpected problems in accordance with Institutional Review Board (IRB) procedures and all federal requirements. Collection, recording and reporting of data will be accurate and will ensure the privacy, health and welfare of research participants during and after the study. Participants have the right to withdraw from the study at any time and for any reason, and all participants are made aware that withdrawal will not affect their routine care.

Consent for publication is not applicable as there are no identifying images of other personal details of participants presented.

Data sharing statement

Any data required to support the protocol will be supplied on request. All information obtained as a result of this study or during the conduct of this study will be regarded as confidential. The requirements concerning dissemination of the information derived from this clinical trial are described in the Clinical Trial Agreement.

Open access

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. *Sleep Med. Rev.* 6(2), 97–111 (2003).
2. Edinger JD, Arnedt JT, Bertisch SM *et al.* Behavioral and psychological treatments for chronic insomnia disorder in adults: an American Academy of Sleep Medicine clinical practice guideline. *J. Clin. Sleep Med.* 17(2), 255–262 (2021).

•• **Updated guidelines.**

3. Edinger JD, Arnedt JT, Bertisch SM *et al.* Behavioral and psychological treatments for chronic insomnia disorder in adults: an American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. *J. Clin. Sleep Med.* 17(2), 263–298 (2021).

•• **Evidence base for updated guidelines.**

4. Daley M, Morin CM, LeBlanc M, Gregoire JP, Savard J. The economic burden of insomnia: direct and indirect costs for individuals with insomnia syndrome, insomnia symptoms, and good sleepers. *Sleep* 32(1), 55–64 (2009).
5. Taylor DJ, Mallory LJ, Lichstein KL, Durrence HH, Riedel BW, Bush AJ. Comorbidity of chronic insomnia with medical problems. *Sleep* 30(2), 213–218 (2007).
6. Ozminkowski RJ, Wang S, Walsh JK. The direct and indirect costs of untreated insomnia in adults in the United States. *Sleep* 30(3), 263–273 (2007).

• **Health economics data.**

7. Morgenthaler T, Kramer M, Alessi C *et al.* Practice parameters for the psychological and behavioral treatment of insomnia: an update. An American Academy of sleep medicine report. *Sleep* 29(11), 1415–1419 (2006).
8. Sateia MJ, Buysse DJ, Krystal AD, Neubauer DN, Heald JL. Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: An American Academy of sleep medicine clinical practice guideline. *J. Clin. Sleep Med.* 13(2), 307–349 (2017).
9. Qaseem A, Kansagara D, Forciea MA, Cooke M, Denberg TD. Management of chronic insomnia disorder in adults: a clinical practice guideline from the American College of Physicians. *Ann. Intern. Med.* 165(2), 125–133 (2016).

•• **Clinical practice guideline.**

10. Schutte-Rodin S, Broch L, Buysse D, Dorsey C, Sateia M. Clinical guideline for the evaluation and management of chronic insomnia in adults. *J. Clin. Sleep Med.* 4(5), 487–504 (2008).

11. American Academy of Sleep Medicine. FDA requires new warning for several insomnia medications (2020). https://aasm.org/fda-new-warning-insomnia-medications/).

12. Food & Drug Administration. FDA adds boxed warning for risk of serious injuries caused by sleepwalking with certain prescription insomnia medicines (2020). www.fda.gov/drugs/drug-safety-and-availability/fda-adds-boxed-warning-risk-serious-injuries-caused-sleepwalking-certain-prescription-insomnia

13. Wu JQ, Appleman ER, Salazar RD, Ong JC. Cognitive behavioral therapy for insomnia comorbid with psychiatric and medical conditions: a meta-analysis. *JAMA Intern. Med.* 175(9), 1461–1472 (2015).

14. Espie CA, Emsley R, Kyle SD *et al.* Effect of digital cognitive behavioral therapy for insomnia on health, psychological well-being, and sleep-related quality of life: a randomized clinical trial. *JAMA Psychiatry* 76(1), 21–30 (2019).

15. Ancoli-Israel S, Lieberman JA 3rd. Insomnia in primary care: overcoming diagnostic and treatment barriers. Introduction. *Postgrad. Med.* 116(Suppl. 6), 4–6 (2004).

16. Thomas A, Grandner M, Nowakowski S, Nesom G, Corbitt C, Perlis ML. Where are the behavioral sleep medicine providers and where are they needed? a geographic assessment. *Behav. Sleep Med.* 14(6), 687–698 (2016).

• **Key data on geographic distribution of cognitive behavioral therapy for insomnia providers.**

17. MedCity News. Pear gets FDA clearance for insomnia therapeutic (2020). https://medcitynews.com/2020/03/pear-therapeutics-gets-fda-clearance-for-insomnia-therapeutic/

18. Food & Drug Administration. Section 510(k) approval letter for Somryst, computerized behavioral therapy device for pschiatric disorders, Class II (2020). www.accessdata.fda.gov/cdrh_docs/pdf19/K191716.pdf

19. Ritterband LM, Thorndike FP, Ingersoll KS *et al.* Effect of a web-based cognitive behavior therapy for insomnia intervention with 1-year follow-up: a randomized clinical trial. *JAMA Psychiatry* 74(1), 68–75 (2016).

•• **Pivotal trial.**

20. Christensen H, Batterham PJ, Gosling JA *et al.* Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. *Lancet Psychiatry* 3(4), 333–341 (2016).

•• **Pivotal trial.**

21. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - a systematic review and meta-analysis of randomized controlled trials. *Sleep Med. Rev.* 30, 1–10 (2015).

22. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (*5th Edition*). American Psychiatric Association, DC, USA (2013).

23. Morin CM, Belleville G, Belanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep* 34(5), 601–608 (2011).

24. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. *J. Affect. Disord.* 114(1), 163–173 (2009).

25. Jordan P, Shedden-Mora MC, Lowe B. Psychometric analysis of the Generalized Anxiety Disorder scale (GAD-7) in primary care using modern item response theory. *PLoS ONE* 12(8), e0182162 (2017).

26. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep* 15(4), 376–381 (1992).

27. Gandek B, Ware JE, Aaronson NK *et al.* Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. *J. Clin. Epidemiol.* 51(11), 1171–1178 (1998).

28. Dworkin SL. Sample size policy for qualitative studies using in-depth interviews. *Arch. Sex. Behav.* 41(6), 1319–1320 (2012).

29. Muller M, Kogan S. Grounded theory method in human-computer interaction (HCI) and computer-supported cooperative work (CSCW) (2021). www.researchgate.net/profile/Michael_Muller/publication/265314097_Grounded_Theory_Method_in_HCI_and_CSCW/links/54dac1a70cf233119bc3e511/Grounded-Theory-Method-in-HCI-and-CSCW.pdf

30. Dimond JP, Fiesler C, DiSalvo B, Pelc J. Qualitative data collection technologies: a comparison of instant messaging, email, and phone (2021). https://dl.acm.org/doi/abs/10.1145/2389176.2389218

31. Furniss D, Blandford A, Curzon P. Confessions from a grounded theory PhD: experiences and lessons learnt. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* DOI: https://doi.org/10.1145/1978942.1978960 113–122 (2011).

Systematic Review

For reprint orders, please contact: reprints@futuremedicine.com

Journal of Comparative Effectiveness Research

A systematic review of noninferiority margins in oncology clinical trials

Mahmoud Hashim[1], Talitha Vincken[1], Florint Kroi[1], Samron Gebregergish[1], Mike Spencer[2], Jianping Wang[2], Tobias Kampfenkel[3], Annette Lam[2] & Jianming He*[,2]

[1]Ingress-Health, Rotterdam, NJ 3012, Netherlands

[2]Janssen Global Services, LLC, Raritan, NJ 08869, USA

[3]Janssen Research & Development, LLC, CB Leiden 2333, Netherlands

*Author for correspondence: Tel.: +1 908 927 2566; jhe2@its.jnj.com

Aim: A systematic literature review was conducted to identify and characterize noninferiority margins for relevant end points in oncology clinical trials. **Materials & methods:** Randomized, controlled, noninferiority trials of patients with cancer were identified in PubMed and Embase. **Results:** Of 2284 publications identified, 285 oncology noninferiority clinical trials were analyzed. The median noninferiority margin was a hazard ratio of 1.29 (mean: 1.32; range: 1.05–2.05) for studies that reported time-to-event end points (n = 192). The median noninferiority margin was 13.0% (mean: 12.7%; range: 5.0–20.0%) for studies that reported response end points as absolute rate differences (n = 31). **Conclusion:** Although there was consistency in the noninferiority margins' scale, variability was evident in noninferiority margins across trials. Increased transparency may improve consistency in noninferiority margin application in oncology clinical trials.

First draft submitted: 14 September 2020; Accepted for publication: 29 January 2021; Published online: 17 March 2021

Keywords: cancer • noninferiority clinical trial • noninferiority margin • oncology • randomized controlled trial • systematic literature review

Noninferiority clinical trials are designed to evaluate whether the efficacy of an experimental intervention is not unacceptably worse than that of a standard of care treatment; these types of studies are useful when a new agent is anticipated to have similar efficacy versus a comparator but improved tolerability, a more convenient dosing/administration schedule and/or reduced costs [1,2]. Regulatory agencies, including the US FDA [3] and the EMA [4], have issued guidelines for study design and statistical considerations in noninferiority clinical trials. These studies typically require a large sample size, and consequently, substantial time and resources.

The noninferiority hypothesis is tested by ruling out a prespecified noninferiority margin, defined as the minimum threshold beyond which the experimental intervention is unacceptably worse than the active comparator [1]. When the whole confidence interval (CI) for the primary end point falls within the margin of noninferiority, the null hypothesis is rejected, and the study is considered positive, in other words, noninferiority cannot be disproven. Conversely, when the value of the low boundary of the CI of the primary end point result falls outside of this range, inferiority cannot be disproven. Hence, the selection of this margin is crucial for both the sample size calculation and later interpretation of results [5].

Despite strict methodological and statistical principles governing noninferiority studies, guidance on how to define specific noninferiority margins for different end points is limited. Regulatory agencies recommend that noninferiority margins be based on statistical considerations and clinical judgment, including historical evidence from previous clinical trials [3,4]. In theory, the size of the prespecified noninferiority margin is dependent on several factors, including disease, severity of toxicity and invasiveness relative to the degree of benefit from the control [6]. Several systematic literature reviews have been conducted to assess noninferiority margins used in clinical trials [7–12]. For oncology noninferiority clinical trials in particular, earlier systematic literature reviews focused on the design and quality [7–9,13]; however, results were not presented separately per specific end point. A systematic literature review published in 2019 evaluated only oncology noninferiority clinical trials with overall survival (OS) as a

Future Medicine

primary or coprimary end point [12]. The aim of the current study was to identify previously used noninferiority margins for various relevant end points in oncology noninferiority clinical trials and to explore factors that drive the selection of noninferiority margins.

Materials & methods

Search strategy

We conducted electronic searches of PubMed and Embase on 4 September 2019 for randomized controlled trials (RCTs) with noninferiority study designs for patients with cancer (detailed search terms are shown in Supplementary Table 1); a manual screening of references of included publications was also conducted. The search was limited to English language publications after 1 January 2000. Editorials and letters were excluded from the PubMed search, and editorials, errata, letters, notes, reviews and short surveys were excluded from the Embase search.

Screening & eligibility criteria

Using predefined eligibility criteria, two investigators (T Vincken and F Kroi) reviewed the titles and abstracts of retrieved articles sequentially for inclusion in the analysis. Publications were included if they were based on randomized noninferiority clinical trials of active treatments (i.e., surgical intervention, radiotherapy, adjuvant or neoadjuvant therapy, or systemic treatment) in patients with any type and stage of cancer; conference abstracts and study protocols were considered if they included sufficient details or if it was possible to retrieve supplementary sources associated with the presented study. Publications were excluded if a noninferiority margin was not prespecified. Duplicate publications, in other words, articles identified both in PubMed and Embase were removed. Subsequently, three investigators (T Vincken, F Kroi and S Gebregergish) reviewed the full text of the selected publications for the eligibility criteria described above as well as for relevant efficacy outcomes (OS, other time-to-event end points or response) or safety outcomes; studies that reported only duration of response or adverse events, quality of life or pharmacokinetic/pharmacodynamic outcomes were excluded. Study ID was used to avoid including multiple publications that reported on the same study. Disagreement regarding eligibility between investigators was resolved by consulting with a fourth investigator (M Hashim).

Data extraction

The following attributes were extracted from the full-length publications: ClinicalTrials.gov registry number, trial phase, country (multiple countries [≥2] vs single country), masking (blinded vs open-label), control arm (active vs placebo), number of treatment arms, sample size, age of participants (adult vs pediatric), cancer setting (early-stage vs advanced/metastatic), cancer type (solid tumor vs blood cancer), treatment modality, primary analysis population (intention-to-treat vs per treatment), primary end point, noninferiority margin scale, rationale for the prespecified noninferiority margin and results of the noninferiority test (successful vs failed; Supplementary Table 2). The dataset was validated based on independent extraction by three investigators (T Vincken, F Kroi and S Gebregergish).

Statistical analysis

Noninferiority margins in the different trials were described on absolute or relative scales. On an absolute scale, the noninferiority margin was expressed as the absolute difference between values of the two treatment groups, and the unit of the noninferiority margin was the same as the unit of the outcome. On a relative scale, the noninferiority margin was expressed as a ratio that compared the two treatment groups, for example, an hazard ratio (HR) between the two treatment groups, with the study outcome measured as the time-to-event. Trials with noninferiority margins reported as HRs for time-to-event outcomes and the absolute rate difference reported for response outcomes were considered for further analysis (primary outcome of this review). For other scales and end points, only descriptive statistics are presented.

The relationships between the reported noninferiority margin and prespecified trial and population characteristics were evaluated with both simple and multiple linear regression models fitted with the noninferiority margin as a dependent variable, available study characteristics and sample size as independent variables. For those variables with statistically significant coefficients ($p < 0.05$) in the multiple model, violin plots were used to visualize the distribution of noninferiority margins.

The following variables were considered: three most common cancers (colorectal, breast or lung cancer) in the studies reviewed, studies published in 2006 or prior (to coincide with EMA guidance [4] and initial Consolidated

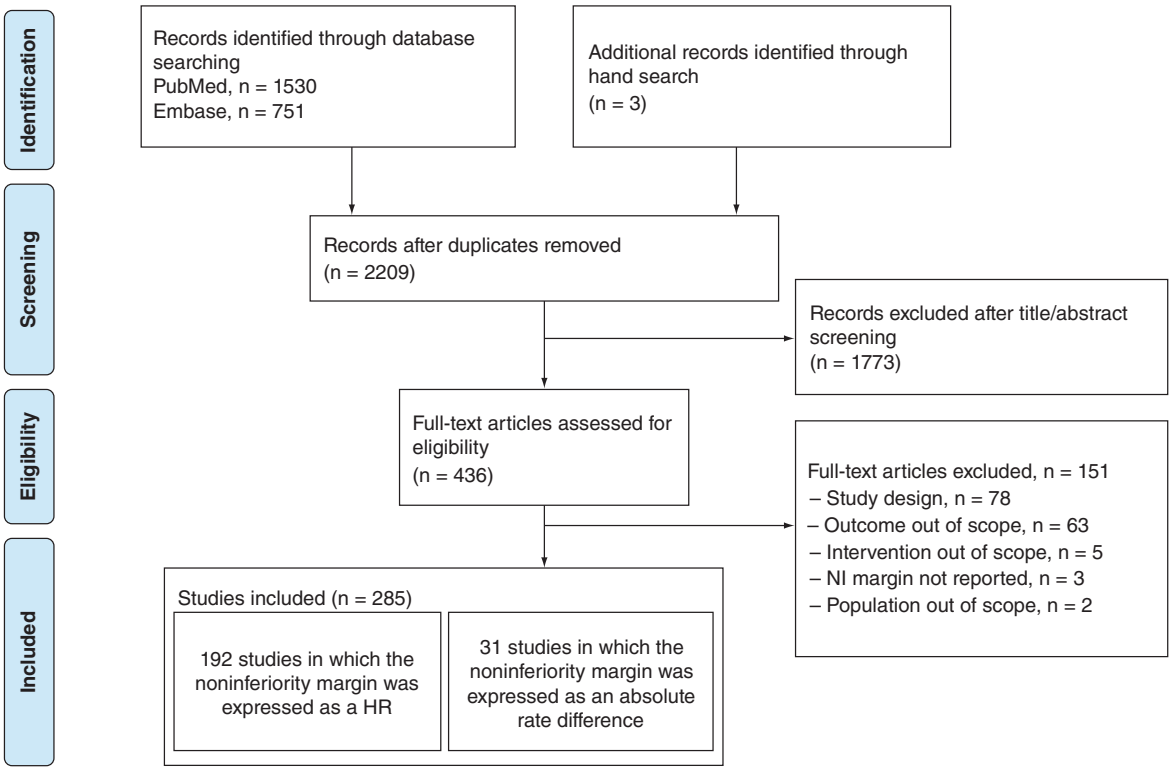


Figure 1. PRISMA flow diagram.
HR: Hazard ratio; NI: Noninferiority.

Standards of Reporting Trials [CONSORT] guidelines for noninferiority trials [14]); studies published from 2007 to 2010 (to coincide with draft FDA guidance [15] and updated CONSORT guidelines for noninferiority trials [16]); studies published from 2011 to 2016 (to coincide with final FDA guidance for noninferiority trials [3]); and studies published in or after 2017; Phase III or IV versus Phase II or not reported; multiple countries versus single country recruitment; open-label, blinded or not reported; active- versus placebo-controlled; adult versus pediatric or mixed-age populations; advanced/metastatic versus early-stage disease; solid tumor versus blood cancer; intention-to-treat versus other analyses; OS versus other primary end point for time-to-event outcomes and overall response rate (ORR) versus other primary end point for response outcomes; and rationale for the noninferiority margin specified versus not specified. Since treatment modality is highly correlated with cancer setting (advanced/metastatic vs early-stage disease), it was not included as a variable.

Results

Selection of publications

A total of 2284 publications were identified in PubMed (n = 1530), Embase (n = 751) and by additional manual screening of the included reports (n = 3). Of 2209 publications that remained after removal of duplicates, 436 records qualified for full-text screening; of these, 285 noninferiority clinical trials met the eligibility criteria and were included in the analysis (Figure 1). A bibliography of the included articles is provided in the Supplementary Material.

Noninferiority clinical trial characteristics

The noninferiority clinical trials included in this systematic review were most commonly Phase III (76.8%), open-label (74.7%), conducted in multiple countries (64.6%) and had active control arms (98.6%; Table 1). The three most common cancers included in the trials were colorectal (22.9%), breast (19.3%) and lung cancer (14.6%). The study populations consisted primarily of adults (97.2%), patients with advanced or metastatic disease (57.5%) and patients with solid tumors (86.0%). In the majority of studies, noninferiority assessment was carried out in the intention-to-treat population (74.0%). The rationale for the prespecified noninferiority margin was reported for

Table 1. Characteristics of included noninferiority clinical trials.

Variable	n (%)
Total	285 (100.0)
Publication date	
2006 or prior	24 (8.4)
2007–2010	41 (14.4)
2011–2016	140 (49.1)
2017 or after	80 (28.1)
Trial phase	
Phase II	27 (9.5)
Phase III	219 (76.8)
Phase IV	3 (1.1)
Not reported	36 (12.6)
Continent	
Asia	97 (34)
Europe	89 (31.2)
Multiple continents	83 (29.1)
North America	13 (4.6)
Australia	1 (0.4)
Africa	1 (0.4)
South America	1 (0.4)
Country	
Multiple Countries	184 (64.6)
Single Country	101 (35.4)
Masking	
Open-label	213 (74.7)
Blinded	22 (7.7)
Not reported	50 (17.5)
Control arm	
Active	281 (98.6)
Placebo	4 (1.4)
No. of treatment arms	
2	265 (93)
3	17 (6.0)
4	3 (1.1)
Total sample size	
≤1000 patients	217 (76.1)
>1000 patients	68 (23.9)
Age of participants	
Adult	277 (97.2)
Pediatric	7 (2.5)
Mixed	1 (0.4)
Cancer setting	
Advanced/metastatic	164 (57.5)
Early-stage	121 (42.5)
Cancer type	
Solid tumor	245 (86.0)
Blood cancer	40 (14.0)

[†]Included very good partial response, complete response, no progressive disease, progressive disease and sonographic recurrence.
[‡]Included expert opinion, assumptions and noninferiority margin in previous studies.
[§]Publications based on a study protocol.
DFS: Disease-free survival; ORR: Overall response rate; OS: Overall survival; PFS: Progression-free survival; RFS: Relapse-free survival; TTP: Time to progression; TTR: Time to relapse.

Table 1. Characteristics of included noninferiority clinical trials (cont.).

Variable	n (%)
Top three cancers by site	
Colorectal cancer	44 (22.9)
Breast cancer	37 (19.3)
Lung cancer	28 (14.6)
Treatment modality	
Systemic	198 (69.5)
Surgical	27 (9.5)
Adjuvant/Neoadjuvant	20 (7.0)
Radiotherapy	20 (7.0)
Combination	20 (7.0)
Primary analysis population	
Intention-to-treat analysis	211 (74.0)
Per protocol analysis	44 (15.4)
Not reported	30 (10.5)
Primary end point	
OS	96 (33.7)
DFS, RFS, TTR or local recurrence	74 (26.0)
PFS or TTP	70 (24.6)
ORR	20 (7.0)
Biochemical or clinical failure	7 (2.5)
Safety	2 (0.7)
Other†	16 (5.6)
Noninferiority margin scale	
Time-to-event end points	194 (68.1)
Hazard ratio	192 (67.4)
Absolute scale	2 (0.7)
Binary end points	91 (31.9)
Absolute scale	88 (30.9)
Absolute difference in survival rate	47 (16.5)
Absolute difference in %, response end points	31 (10.9)
Absolute difference in %, other efficacy end points	8 (2.8)
Absolute difference in %, safety end points	2 (0.7)
Relative scale	3 (1.1)
Rationale for prespecified noninferiority margin	
Based on historical survival rates	111 (38.9)
Based on effect size of active over control treatment in prior trials	92 (32.3)
Statistically appropriate / feasible	25 (8.8)
Other reasons‡	22 (7.7)
Not reported	35 (12.3)
Results of the noninferiority test	
Successful	174 (61.1)
Failed	84 (29.5)
Unknown§	27 (9.5)

†Included very good partial response, complete response, no progressive disease, progressive disease and sonographic recurrence.
‡Included expert opinion, assumptions and noninferiority margin in previous studies.
§Publications based on a study protocol.
DFS: Disease-free survival; ORR: Overall response rate; OS: Overall survival; PFS: Progression-free survival; RFS: Relapse-free survival; TTP: Time to progression; TTR: Time to relapse.

most (87.7%) of the included studies; estimation was based on historical data (survival rates or effect size of active over control treatment in prior studies) in 71.2% of the trials. Nearly a third (29.5%) of the included trials failed to establish noninferiority for the active treatment over the control treatment. A small proportion of the studies used a placebo control arm (1.4%) or investigated a pediatric patient population (2.5%); therefore, these variables were excluded from the regression models. The mean sample size in the included trials was 829 (range: 7–10,270) and the median (25th–75th percentile) sample size was 509 (272–981).

In all, 194 (68.1%) studies used noninferiority margins for time-to-event end points (Table 1); 192 studies in which the noninferiority margin was expressed as an HR between treatment groups were selected for further analysis, as per the prespecified primary outcome of this review. A total of 91 (31.9%) studies used noninferiority margins to evaluate binary end points; 31 studies that expressed the absolute difference in rates in treatment groups for response end points were included in further analysis.

Noninferiority margins for time-to-event end points

In the 192 studies that reported noninferiority margins for time-to-event end points as an HR, the mean/median sample size was 990/618 patients (range: 21–10,273); 134 studies (69.8%) evaluated populations of ≤1000 patients, with a mean/median sample size of 490/472 (range: 21–999). Across all trials, noninferiority margins ranged from 1.05 to 2.05 with mean and median values of 1.32 and 1.29, respectively; corresponding values for studies of ≤1000 patients were 1.33 and 1.30. As shown in Figure 2A, larger sample size was associated with lower noninferiority margins for time-to-event end points. Visual examination of the scatter plots showed half a symmetrical inverted funnel above the no effect estimate (i.e., HR: 1); with sample size used as a measure of precision, these results indicated that the risk of publication bias was unlikely.

Summary statistics for noninferiority margins of time-to-event end points are shown in Table 2; trends for higher noninferiority margins in study populations of ≤1000 patients appeared to be consistent with the overall analysis of time-to-event end points. The mean and median noninferiority margins, respectively, for the three most common cancers were, 1.31 and 1.28 for colorectal, 1.31 and 1.29 for breast and 1.25 and 1.25 for lung cancer. Eighty studies reported noninferiority margins as an HR for OS (mean/median sample size: 700/606 patients; range: 21–2135) with mean and median noninferiority margins of 1.28 and 1.25, respectively, in all studies. Fifty-seven studies reported noninferiority margins as an HR for progression-free survival (PFS) or time-to-progression (TTP) (mean/median sample size: 548/450 patients; range: 58–2126) with mean and median noninferiority margins of 1.30 and 1.25, respectively, in all studies. The range of noninferiority margins was smaller for OS (1.05–1.82) and PFS (1.08–1.80) compared with time-to-event end points overall (1.05–2.05). In an analysis of trials of patients who received chemotherapy (n = 83) versus targeted therapy (n = 55), mean and median noninferiority margins were similar for the systemic treatments (data not shown).

Based on the simple linear regression models, trial phase (Phase III/IV vs Phase II or not reported), patient recruitment (multiple countries vs single country), cancer setting (advanced/metastatic vs early disease) and primary end point (OS vs others) were significantly associated with lower noninferiority margins; based on the multiple model, trial phase and cancer setting remained significantly associated with lower noninferiority margins (Supplementary Tables 3 & 4). Additionally, in the multiple model, studies with lower sample size and those reporting rationale behind margin specification reported lower margins. Violin plots of noninferiority margin scales for HRs according to these variables are shown in Supplementary Figure 1.

For two studies that used the absolute difference in median survival times (sample sizes: 271 and 284 patients), the noninferiority margin was assumed to be 1.5 months in each study. In both studies, this was on the basis of survival rates that come from unspecified sources.

Noninferiority margins for binary end points

Among 31 studies that reported noninferiority margins for response end points as absolute rate difference, the mean/median sample size was 289/212 patients (range: 7–1229). Noninferiority margins ranged from 5.0 to 20.0%, with mean and median values of 12.7 and 13.0%, respectively. Similar to the trend observed for time-to-event end points, larger sample size was associated with lower noninferiority margins for response end points, with half a symmetrical inverted funnel above the no effect estimate (i.e., absolute difference = 0%) indicating absence of bias (Figure 2B).

Summary statistics for response end points and prespecified subgroups are shown in Table 3. A total of 19 (61.3%) of these studies (mean/median sample size: 246/212 patients; range: 7–719) evaluated ORR as the primary end

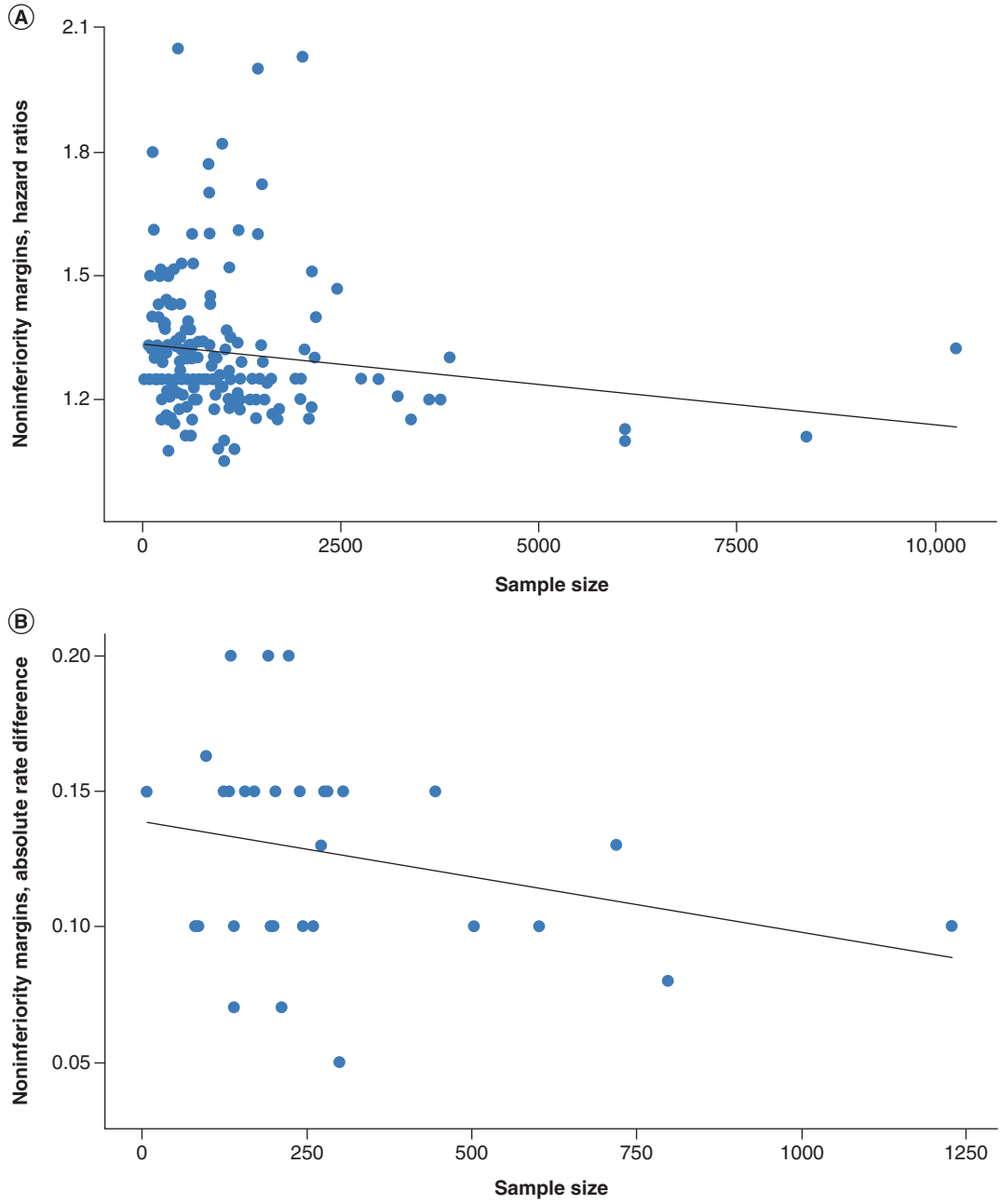


Figure 2. Distribution of noninferiority margins according to trial sample size for (A) time-to-event end points using hazard ratio as the scale and (B) response end points using the absolute rate difference as the scale.

point; in this group, the mean noninferiority margin was 13.5% and median was 15.0%. The remaining 12 (38.7%) studies (mean/median sample size: 358/211; range: 81–1229) assessed other response end points; mean and median noninferiority margins were 11.5 and 10.0%, respectively. Mean and median noninferiority margins were similar for trials of chemotherapy (n = 17) and targeted therapy (n = 7; data not shown). No variables were predictive of the noninferiority margin, based on the simple or multiple linear regression model (Supplementary Tables 3 & 4).

For 47 studies that assessed the absolute difference in survival rate as the noninferiority margin scale (mean/median sample size: 482/360 patients; range: 70–2073), mean and median noninferiority margins were 11 and 10%, respectively. For eight studies (mean/median sample size: 1512/1374; range: 89–2090) that reported absolute difference in percentage for efficacy end points other than response, for example, locoregional recurrence

Table 2. Summary statistics for time-to-event end points using hazard ratio as noninferiority margin scale in all studies and prespecified subgroups.

	Hazard ratio							
	n	Mean	STD	Minimum	25th percentile	Median	75th percentile	Maximum
All studies	192	1.32	0.16	1.05	1.22	1.29	1.35	2.05
Colorectal cancer	44	1.31	0.13	1.10	1.24	1.28	1.33	1.82
Breast cancer	37	1.31	0.16	1.11	1.23	1.29	1.33	2.03
Lung cancer	28	1.25	0.08	1.11	1.18	1.25	1.32	1.50
Trials by sample size								
Trials with >1000 patients	58	1.29	0.19	1.05	1.19	1.25	1.32	2.03
Trials with ≤1000 patients	134	1.33	0.15	1.08	1.25	1.30	1.37	2.05
Primary end point								
OS	80	1.28	0.13	1.05	1.20	1.25	1.33	1.82
PFS or TTP	57	1.30	0.13	1.08	1.22	1.25	1.34	1.80
DFS, RFS, TTR or local recurrence	48	1.37	0.22	1.10	1.25	1.31	1.46	2.05
Biochemical or clinical failure	7	1.42	0.19	1.21	1.25	1.37	1.61	1.72
Publication date								
2006 or prior	13	1.26	0.09	1.11	1.25	1.25	1.26	1.51
2007 to 2010	31	1.26	0.09	1.08	1.18	1.25	1.33	1.50
2011 to 2016	100	1.33	0.17	1.05	1.21	1.30	1.42	2.00
After 2017	48	1.34	0.19	1.08	1.25	1.31	1.37	2.05
Trial phase								
Phase III or IV	160	1.29	0.13	1.05	1.21	1.25	1.33	2.03
Phase II or not reported	32	1.45	0.23	1.15	1.26	1.39	1.51	2.05
Country								
Multiple countries	79	1.28	0.17	1.05	1.20	1.25	1.33	2.00
Single country	113	1.34	0.15	1.10	1.25	1.32	1.39	2.05
Masking								
Open-label	110	1.32	0.16	1.10	1.25	1.30	1.34	2.05
Blinded	14	1.28	0.12	1.08	1.24	1.25	1.32	1.60
Cancer setting								
Advanced/metastatic	118	1.28	0.12	1.05	1.20	1.25	1.33	1.82
Early-stage	74	1.38	0.20	1.10	1.25	1.32	1.50	2.05
Cancer type								
Solid tumor	176	1.31	0.16	1.05	1.23	1.27	1.34	2.05
Blood cancer	16	1.39	0.20	1.11	1.21	1.36	1.51	1.77
Primary analysis								
Intention-to-treat	144	1.31	0.16	1.05	1.22	1.26	1.36	2.03
Other	48	1.33	0.18	1.08	1.22	1.30	1.34	2.05
Rationale for noninferiority margin								
Specified	171	1.31	0.16	1.05	1.22	1.26	1.33	2.05
Not specified	21	1.37	0.16	1.11	1.25	1.34	1.47	1.80

DFS: Disease-free survival; OS: Overall survival; PFS: Progression-free survival; RFS: Relapse-free survival; TTP: Time to progression; TTR: Time to relapse. STD: Standard deviation.

and sonographic recurrence, the mean and median noninferiority margins were 7.5 and 8.0%, respectively. For two studies that assessed safety end points with the absolute percentage difference as the noninferiority margin scale (sample sizes: 200 and 217 patients), noninferiority margins were 10 and 15%, respectively. For three studies that assessed relative risk for binary efficacy outcomes as the noninferiority margin scale (sample sizes: 447, 707 and 501 patients), noninferiority margins were reported to be 1.14, 1.25 and 1.15, respectively.

Table 3. Summary statistics for response end points using the absolute rate difference as noninferiority margin scale in all studies and prespecified subgroups.

		Percentage point (%)						
	n	Mean	STD	Minimum	25th percentile	Median	75th percentile	Maximum
All studies	31	12.7	3.9	5.0	10.0	13.0	15.0	20.0
Colorectal cancer	2	12.5	3.5	10.0	10.0	12.5	15.0	15.0
Breast cancer	5	11.4	3.5	7.0	8.5	10.0	15.0	15.0
Lung cancer	6	14.7	3.3	10.0	12.3	15.0	16.3	20.0
Primary response end point								
ORR	19	13.5	3.7	7.0	10.0	15.0	15.0	20.0
Other response end points	12	11.5	4.0	5.0	10.0	10.0	15.0	20.0
Publication date								
2006 or prior	4	13.3	5.4	8.0	8.5	12.5	18.8	20.0
2007 to 2010	3	13.3	2.9	10.0	10.0	15.0	15.0	15.0
2011 to 2016	13	12.7	3.9	5.0	10.0	15.0	15.0	20.0
After 2017	11	12.4	4.0	7.0	10.0	13.0	15.0	20.0
Trial phase								
Phase III or IV	21	12.6	4.1	7.0	10.0	10.0	15.0	20.0
Phase II or not reported	10	12.9	3.5	5.0	10.0	15.0	15.0	16.3
Country								
Multiple countries	25	12.5	4.2	5.0	10.0	10.0	15.0	20.0
Single country	6	13.5	2.0	10.0	12.3	14.0	15.0	15.0
Masking								
Open-label	18	13.0	4.1	5.0	10.0	15.0	15.0	20.0
Blinded	5	11.6	3.1	7.0	8.5	13.0	14.0	15.0
Cancer setting								
Advanced/metastatic	18	12.3	4.0	5.0	9.5	14.0	15.0	20.0
Early-stage	13	13.3	3.7	10.0	10.0	13.0	15.0	20.0
Cancer type								
Solid tumor	21	12.8	3.3	7.0	10.0	15.0	15.0	20.0
Blood cancer	10	12.5	5.1	5.0	9.3	11.5	16.3	20.0
Primary analysis								
Intention-to-treat	23	13.4	3.8	7.0	10.0	15.0	15.0	20.0
Other	8	10.9	3.8	5.0	7.8	10.0	15.0	15.0
Rationale for noninferiority margin								
Specified	24	13.2	3.7	7.0	10.0	14.0	15.0	20.0
Not specified	7	11.2	4.4	5.0	7.0	10.0	15.0	16.3

ORR: Overall response rate; STD: Standard deviation.

Discussion

This study aimed to identify and characterize previously applied noninferiority margins for relevant end points in oncology noninferiority clinical trials and included a variety of cancer types, settings and treatments. Across 192 trials reporting time-to-event end points with HRs, mean and median noninferiority margins were 1.32 and 1.29, respectively. Across 31 trials reporting response end points with absolute rate difference, mean and median noninferiority margins were 12.7 and 13.0%, respectively. There was substantial variation in noninferiority margins for both time-to-event end points (range: 1.05–2.05) and response end points (range: 5.0–20.0%).

Noninferiority margins in the literature

The median noninferiority margin values reported here are in line with those described in a prior systematic literature review of oncology noninferiority clinical trials published from January 2001 to January 2011, in which the median noninferiority margin was 1.25 for 34 studies reporting time-to-event end points and 12.5% for 28 studies reporting binary end points [8]. In that analysis, the noninferiority margin range for time-to-event outcomes

(1.10–1.50) was narrower than that observed here, whereas the range for binary end points (4–25%) was slightly broader [8]. A list of the included oncology noninferiority clinical trials was not published with that systematic literature review [8]; thus, it is not possible to directly compare our findings with the results of the earlier analysis. A systematic literature review (conducted in March 2018 with no date limitations) that evaluated noninferiority criteria for HRs for 23 oncology noninferiority clinical trials with OS as a primary or coprimary end point reported a range of 1.08–1.33 [12], which was also narrower than the range for OS in our analysis (1.05–1.82). This may be explained, in part, by the fact that the trials included in the analysis by Gyawali *et al.* were all Phase III studies and evaluated patients with solid tumors [12]; these variables were significantly associated with lower noninferiority margins for time-to-event end points in our analysis.

Selection of noninferiority margin

Guidelines from the FDA and EMA recommend that both statistical and clinical judgment be used for the selection of a noninferiority margin [3,4]. Statistical reasoning should be based on historical data for the active comparator, preferably with the noninferiority margin defined according to pooled effect estimates from multiple prior RCTs and clinical judgment used to establish the proportion of the known effect of the active control versus placebo that must be maintained with the experimental agent [2]. However, practical considerations for trial feasibility, such as sample size, must be weighed against the clinical relevance of the noninferiority threshold and may, in part, drive differences in trial design and margin specification observed here. In simple and multiple linear regression models, the timing of study publication relative to EMA guidance (from 2007 to 2010 vs in 2006 or prior), draft FDA guidance (from 2011 to 2016 vs 2007 to 2010) or final FDA guidance (on or after 2017 vs 2011 to 2016), was not predictive of noninferiority margin. This suggests that the practice of noninferiority margin selection has not changed over time, but it is not readily apparent if it is consistent with regulatory guidance.

In general, there was considerable consistency in the scale used for noninferiority margins: most time-to-event end points were described with HRs and most binary end points were described using absolute difference in percentage point. However, there was a substantial variation in the prespecified margins and the rationale for choosing those margins. Since the benefit–risk assessment for cancer treatments differs from other therapeutic areas, special considerations should be taken when designing oncology noninferiority clinical trials, with particular attention to regulatory guidance on prespecification of the noninferiority margin, sample size and analysis population [3,4]. Increased transparency regarding the methods for specification of noninferiority margins will aid in design of future trials. Researchers will have a better understanding of the methodological issues and challenges involved in selection of noninferiority margins based on previous studies, which in turn can help them address any queries while designing future studies [2,17]. Transparency will also facilitate comparison between different trials utilizing the same noninferiority margins and could help researchers identify areas that need further study [2].

Rationale for selection of noninferiority margins

Of note, we found that the rationale for the prespecified noninferiority margin was not stated in 12.2% of the analyzed trials, and three studies that were identified in the search had to be excluded because they did not report the selected noninferiority margin. Since the first CONSORT extension for noninferiority trials in 2006, it has been recommended that the noninferiority margin and rationale for its selection be included in publications of randomized noninferiority clinical trials [14,16]. Despite this, earlier systematic literature reviews, published in 2012 and 2013, have also described inadequate reporting compared with CONSORT guidelines [7–9].

We also identified two systematic literature reviews published in the last 5 years of noninferiority trials outside of oncology. The first review explored the noninferiority margins used in vaccine RCTs, and results indicate that of the 143 trials, 66% used a margin of 10, 23% used margins lower than 10% and 11% used margins larger than 10% [18]. The authors therefore conclude that while most noninferiority vaccine RCTs used a noninferiority margin of 10% for difference, the variation in the margins was primarily due to the lack of rationale and unclear guidelines on the selection of noninferiority margins. Similarly, a second review assessing noninferiority margins in anti-infective trials showed that of the 227 trials, only 36.6% had a clear rationale for selection of noninferiority margins and 15% had misleading conclusions [19].

Timing of study publication relative to the issuance of the first CONSORT guidelines (from 2007 to 2010 vs in 2006 or prior) or updated guidelines (from 2011 to 2016 vs from 2007 to 2010) was not predictive of noninferiority margin in our simple and multiple linear regression models. The majority of the studies included in

this analysis used the ITT population; however, FDA guidance for noninferiority studies recommends the use of a per protocol population [3].

Sample size considerations

Sample size planning is a crucial part of any trial design. It is well established that a prespecified noninferiority margin has a direct impact on the size of a noninferiority trial, since if the noninferiority margin is reduced the sample size is increased and *vice versa* [3]. Our finding that Phase III or IV studies were associated with lower noninferiority margins for time-to-event end points versus Phase II studies or studies for which the phase was not reported is likely due to their higher quality statistical planning and larger sample size. These factors may also be applied to the finding that studies conducted in multiple ($\geq$2) countries were associated with lower noninferiority margins than those conducted in a single country. We additionally found that advanced/metastatic cancer (versus early-stage disease) and solid tumors (versus blood cancer) were associated with lower inferiority margins for time-to-event end points. The lower noninferiority margins observed in trials of advanced/metastatic versus early-stage cancer may be attributed to the fact that more events are expected within a shorter follow-up duration in patients with advanced disease, and thus, these studies have more power to show noninferiority for a given sample size. Moreover, smaller differences in efficacy outcomes are expected to be more meaningful in patients at higher risk, with a greater absolute impact that is reflected by the noninferiority margin.

Limitations & areas for future research

There are several limitations associated with this systematic literature review.

First, reporting of the methods and the rationale for specifying noninferiority margins in the included publications was generally short and often ambiguous. Consequently, we were unable to use this information to provide a specific recommendation on how noninferiority margins should be described, defined or justified in noninferiority clinical trials. Second, we analyzed cancer trials in general, for a comprehensive analysis, rather than focusing on a particular oncology setting or indication. Evaluation of noninferiority margins for specific types of cancer or treatments could be an area of future research. Third, we limited our analysis to publications of efficacy and safety outcomes. Health-related quality of life is also a key outcome of oncology trials but is rarely selected as the primary end point and thus represents an evidence gap for future analysis.

In this paper, we have focused on the noninferiority margin values used in previous oncology noninferiority trials rather than exploring the associations between positive results (i.e., a finding of noninferiority between two treatment arms) and characteristics of those studies. This is an interesting topic that warrants further research. Additionally, an assessment of the adequacy of noninferiority margins was not done as part of our analysis due to the retrospective nature of the studies, but could provide valuable insights to future researchers. Finally, other variables, including the effects of study duration, selection of active control, statistical power and alpha on margin choice, were not within the scope of this work and represent other areas for future investigation.

The noninferiority margins for key oncology end points identified here can aid in the interpretation of data from indirect treatment comparisons, for circumstances in which head-to-head trials have not been conducted or are not feasible. An earlier, targeted literature review of 99 publications based on oncology noninferiority clinical trials, identified mean noninferiority margins for PFS and OS as HRs of 1.333 and 1.298, respectively, for studies of $\leq$1000 patients [20], which is consistent with the findings here, based on a larger publication sample size. These noninferiority margins have been applied to matching adjusted indirect comparisons to categorize differences between treatment regimens, with results that did not achieve superiority or inferiority and did not qualify per the noninferiority criteria (HR: 1.333 for PFS and HR: 1.298 for OS) [20] treated as inconclusive [21].

Future perspective

With the development of new cancer treatments, more noninferiority clinical trials can be expected in the near future. Greater number of noninferiority trials will result in a larger dataset that can be utilized for systematic analysis. It will be interesting to see if there is an improvement in reporting of these noninferiority trials compared with the trials reviewed in our analyses. In the future, with a larger dataset and more complete reporting of noninferiority trials, it may be possible to make recommendations for optimal noninferiority margins in oncology clinical trials.

Conclusion

This systematic literature review identified and synthesized previously used noninferiority margins for time-to-event and response end points in randomized, controlled, noninferiority clinical trials of patients with cancer. There was considerable consistency in the scale used for noninferiority margins: most time-to-event end points were described with HRs and most binary end points were described using absolute difference in percentage. There was considerable variation in prespecified noninferiority margins across trials. Greater transparency about the selection of noninferiority margins and further research are needed to improve application and reporting of noninferiority margins in oncology noninferiority clinical trials.

Summary points

- Noninferiority clinical trials are designed to evaluate if the efficacy of an experimental intervention is not unacceptably worse than that of a standard of care treatment.
- A systematic literature review was performed to evaluate previously used noninferiority margins for relevant end points in oncology noninferiority clinical trials.
- Among 192 studies that reported noninferiority margins for time-to-event end points as a hazard ratio, mean and median values were 1.32 and 1.29, respectively, with a range of 1.05 to 2.05.
- Among 31 studies that reported noninferiority margins for response end points as absolute rate difference, mean and median values were 12.7 and 13.0%, respectively, with a range of 5.0–20.0%.
- Increased transparency regarding the specification of noninferiority margins is needed to improve consistency in their definition and application in oncology noninferiority clinical trials.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/cer-2020-0200

Author contributions

M Hashim, T Vincken, F Kroi and S Gebregergish performed the systematic literature review and analysis. M Spencer, J Wang, T Kampfenkel, A Lam and J He designed the analyses. All the authors participated in data interpretation, contributed to drafting of the manuscript and provided final approval for submission.

Financial & competing interests disclosure

These analyses were sponsored by Janssen Research & Development, LLC. M Hashim, T Vincken, F Kroi and S Gebregergish are employees of Ingress-Health, which was hired by Janssen to conduct this research. M Spencer, J Wang, T Kampfenkel, A Lam and J He are employees of Janssen. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Medical writing support was provided by Joanna Bloom, PhD, of Eloquent Scientific Solutions and was funded by Janssen Global Services, LLC.

References

Papers of special note have been highlighted as: •• of considerable interest

1. Fleming TR, Odem-Davis K, Rothmann MD, Li Shen Y. Some essential considerations in the design and conduct of non-inferiority trials. *Clin. Trials* 8(4), 432–439 (2011).
2. Althunian TA, de Boer A, Groenwold RHH, Klungel OH. Defining the noninferiority margin and analysing noninferiority: an overview. *Br. J. Clin. Pharmacol.* 83(8), 1636–1642 (2017).
3. US Food and Drug Administration. *Guidance Document: Non-Inferiority Clinical Trials to Establish Effectiveness.*(2016). www.fda.gov/media/78504/download

•• **US FDA's regulatory document includes guidelines on different aspects of noninferiority trial design.**

4. European Medicines Agency. Guideline on the choice of the non-inferiority margin. [Online] Committee For Medicinal Products For Human Use (CHMP) (2005). www.ema.europa.eu/en/documents/scientific-guideline/guideline-choice-non-inferiority-margin_en.pdf

•• **European Medicines Agency's regulatory document includes guidelines on selection of noninferiority margins.**

5. Jung SH. Design of Phase II non-inferiority trials. *Contemp. Clin. Trials Commun.* 7, 23–27 (2017).

6. *Introduction to Statistical Methods for Clinical Trials.* Cook TD, De Mets DL (Eds). Chapman & Hall/CRC, FL, USA (2008).
7. Tanaka S, Kinjo Y, Kataoka Y, Yoshimura K, Teramukai S. Statistical issues and recommendations for noninferiority trials in oncology: a systematic review. *Clin. Cancer Res.* 18(7), 1837–1847 (2012).

•• **This literature review focused on the design and quality of oncology noninferiority clinical trial, with special emphasis on methodological issues.**

8. Riechelmann RP, Alex A, Cruz L, Bariani GM, Hoff PM. Non-inferiority cancer clinical trials: scope and purposes underlying their design. *Ann. Oncol.* 24(7), 1942–1947 (2013).

•• **This systematic literature review evaluated the purpose underlying selection of noninferiority design and size of the noninferiority margin in oncology trials.**

9. Saad ED, Buyse M. Non-inferiority trials in breast and non-small cell lung cancer: choice of non-inferiority margins and other statistical aspects. *Acta Oncol.* 51(7), 890–896 (2012).

•• **This review assessed the design features of oncology noninferiority clinical trials with regard to the justification of noninferiority margins.**

10. Rehal S, Morris TP, Fielding K, Carpenter JR, Phillips PP. Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open* 6(10), e012594 (2016).
11. Althunian TA, de Boer A, Klungel OH, Insani WN, Groenwold RH. Methods of defining the non-inferiority margin in randomized, double-blind controlled trials: a systematic review. *Trials* 18(1), 107 (2017).
12. Gyawali B, Tessema FA, Jung EH, Kesselheim AS. Assessing the justification, funding, success, and survival outcomes of randomized noninferiority trials of cancer drugs: a systematic review and pooled analysis. *JAMA Netw. Open* 2(8), e199570 (2019).
13. Paesmans M, Grigoriu B, Ocak S *et al.* Systematic qualitative review of randomised trials conducted in nonsmall cell lung cancer with a noninferiority or equivalence design. *Eur. Respir. J.* 45(2), 511–524 (2015).
14. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Group C. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 295(10), 1152–1160 (2006).
15. US Food and Drug Administration. Guidance for industry non-inferiority clinical trials draft guidance (2010). www.fdanews.com/ext/resources/files/archives/n/NoninferiorityGuidance.pdf
16. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. Group C. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 308(24), 2594–2604 (2012).
17. HonOrio HM, Wang L, Rios D. Non-inferiority clinical trials: importance and applications in health sciences. *Braz. Oral Res.* 34(Suppl. 2), e072 (2020).
18. Donken R, de Melker HE, Rots NY, Berbers G, Knol MJ. Comparing vaccines: a systematic review of the use of the non-inferiority margin in vaccine trials. *Vaccine* 33(12), 1426–1432 (2015).
19. Bai AD, Komorowski AS, Lo CKL *et al.* Methodological and reporting quality of non-inferiority randomized controlled trials comparing antibiotic therapies: a systematic review. *Clin. Infect. Dis.* ciaa1353 10.1093/cid/ciaa1353 (2020).
20. Hashim M, He J, Hu P *et al.* Is there a consensus regarding clinically relevant non-inferiority margins used for key oncology endpoints in non-inferiority oncology trials? *Value Health* 21(Suppl. 1), S228 (2018).
21. Mateos MV, San-Miguel J, Goldschmidt H *et al.* The effects of different schedules of bortezomib, melphalan, and prednisone for patients with newly diagnosed multiple myeloma who are transplant ineligible: a matching-adjusted indirect comparison. *Leuk. Lymphoma* 61(3), 680–690 (2020).

Preliminary Communication

Journal of Comparative Effectiveness Research

Optimizing electronic capture of patient-reported outcome measures in oncology clinical trials: lessons learned from a qualitative study

Florence D Mowlem*[,1], Brad Sanderson[1], Jill V Platko[2] & Bill Byrom[1]

[1]Signant Health, London, W6 9RH, UK

[2]Signant Health, Blue Bell, PA 19422, USA

*Author for correspondence: florence.mowlem@signanthealth.com

Aim: To understand the impact of anticancer treatment on oncology patients' ability to use electronic solutions for completing patient-reported outcomes (ePRO). **Materials & methods:** Semi-structured interviews were conducted with seven individuals who had experienced a cancer diagnosis and treatment. **Results:** Participants reported that the following would impact the ability to interact with an ePRO solution: peripheral neuropathy of the hands (4/7), fatigue and/or concentration and memory issues (6/7), where they are in a treatment cycle (5/7). Approaches to improve usability included: larger, well-spaced buttons to deal with finger numbness, the ability to pause a survey and complete at a later point and presenting the recall period with every question to reduce reliance on memory. **Conclusion:** Symptoms associated with cancers and anticancer treatments can impact the use of technologies. The recommendations for optimizing the electronic implementation of patient-reported outcome instruments in this population provides the potential to improve data quality in oncology trials and places patient needs at the forefront to ensure 'fit-for-purpose' solutions.

First draft submitted: 8 July 2020; Accepted for publication: 28 September 2020; Published online: 4 December 2020

Keywords: clinical trials • eCOA • ePRO • ePROMs • oncology • patient-reported outcome measures • usability

Patient-reported outcomes (PROs) are now routinely used as endpoints in oncology clinical trials to elicit the patient perspective, particularly related to symptoms and health-related quality of life (HRQoL) [1] and for supplementing objective clinical endpoints. The importance of utilizing PROs in drug development has also been emphasized by regulatory agencies such as the US FDA [2], and they are increasingly being used in labeling claims for oncology treatments. Between 2012 and 2016, 49 oncology drugs were approved by the FDA and/or EMA for a total of 64 indications; with 45 indications (70.3%) including PRO data in the regulatory submission dossier. Twenty-one of the 45 indications included PRO data on the labeling approved by the EMA [3]. However, both the FDA and EMA identified missing PRO data as problematic for the interpretation of efficacy [3], and the EMA specifically recognizes that collecting PROs from patients with advanced and progressive disease such as cancer, can be more problematic and challenging due to deteriorating health and/or cognitive difficulties [4], which can lead to increased missing data.

With the increasing use of PROs as trial endpoints, comes an increasing drive to capture PROs electronically (ePROs) due to the benefits offered. Electronic collection enables improved data quality due to inbuilt branching (eliminating conflicting entries) and edit checks to eliminate invalid or missing responses, and the ability for data to be reported at defined timepoints outside the clinic environment [5]. When capturing PROs outside of the clinic, electronic capture can offer improved data integrity, including transparent attributability and contemporaneousness due to preprogrammed completion windows, reminders and time/date stamped entries, reducing reporting gaps and enabling more sensitive detection of change. This is particularly relevant in oncology where health status is dynamic throughout treatment and recall of health status is often influenced by current status [6]. Thus, only

Future Medicine

measuring outcomes at site visits that occur at the beginning of a treatment cycle can be misleading; for example, in relation to side effects, this assessment timepoint is likely to occur after recovery from the previous cycle and before the expected time of greatest toxicity of the next cycle of treatment, [7] and therefore not provide an accurate reflection of the patient perspective. Furthermore, drug administration schedules differ (e.g., intravenous [IV] vs oral treatment) and so this can make comparisons between different classes of drugs within one trial challenging. Thus, measuring PROs at home offers an opportunity to capture more comprehensive and accurate patient perspective data.

Furthermore, for studies with at-home PRO assessments (e.g., daily ediaries and surveys/questionnaires) as primary or secondary end points that are going to be used for labeling claims, the FDA request evidence to demonstrate how the timing of measures (or events) specified in the protocol has been enforced [6]. ePRO solutions provide a time-stamped audit trail to corroborate this, offering a solution against the infamous 'parking lot effect' (completion of paper diaries in the parking lot just prior to a clinic visit) [5].

However, the usability of electronic solutions to capture PROs may be influenced uniquely by the impact of the cancer and its treatment. For example, chemotherapy-induced peripheral neuropathy (CIPN) has been reported in the literature, with up to 90% of oncology patients experiencing it at some stage during or after treatment [8,9]. Furthermore, CIPN can persist and some individuals may not recover full dexterity [8,10]. Cancer-related fatigue is also common, can be persistent, and experienced as physical, emotional and/or cognitive tiredness or exhaustion [11]. Cancer Research UK [12] states that it can affect between 25–99% of those undergoing treatment. Individuals can also experience cognitive impairments in multiple domains, including verbal memory, visual spatial skills, attention and psychomotor function (such as fine motor dexterity) [13]. Studies have shown that this is a highly prevalent impact of the disease and treatment, with up to 30% of people experiencing these problems prior to treatment, up to 75% during treatment, and up to 35% experiencing persistent cognitive difficulties post-treatment [14]. Such cognitive impairments resulting from anticancer treatments have been colloquially referred to as 'chemobrain' and more recently 'cancer-associated cognitive change' or 'cancer-treatment-associated cognitive change', as it is not confined to chemotherapy [14–18]. Such symptoms can pose particular challenges for clinical trials where ePROs are completed outside the clinic environment and site-staff are not present to provide guidance and ensure completion.

Given the wealth of benefits offered by electronic measurement, it is imperative to comprehensively understand whether device and software functionality for collecting PROs electronically are 'fit-for-purpose' in this disease population. While there have been studies conducted that look at the feasibility and acceptability of electronic recording of PROs [19], to our knowledge, there has not been a qualitative study in oncology patients exploring how the disease and symptoms associated with its treatment can impact willingness and ability to use technology for recording ePROs. Qualitative research offers a unique opportunity to understand patient experience in greater depth, which is especially pertinent in new and emerging research areas, and findings can be harnessed to inform future quantitative study designs. The objectives of the current study were to conduct semi-structured interviews to: 1) qualitatively investigate the challenges experienced by oncology patients that effect electronic device use for recording PROs; 2) use this qualitative data to inform recommendations for optimising implementation of ePRO instruments in oncology clinical trials.

Materials & methods

This research adheres to the recommendations outlined in the Standards for Reporting Qualitative Research [20].

The interviews detailed in this manuscript were conducted in the context of a broader usability testing project carried out in the UK. The overarching aim of this project was to investigate specific user needs and usability considerations of oncology patients who participate in clinical trials when using Signant Health's electronic Clinical Outcome Assessment (eCOA) software solutions on two devices (a tablet and a provisioned mobile device), to understand how well patients can interact with the software solutions (ediaries, questionnaires and different response scale types) and what problems they experience, subsequently facilitating opportunities to make the solutions more user-friendly for this patient population. Here, we only report on the themes related to how the disease and symptoms associated with its treatment impact technology use. As this project formed a software usability testing evaluation, ethical approval was not required. All participants provided informed consent and signed a confidentiality form.

Sample

Participants were recruited using an external recruitment agency and underwent a telephone screening. To be eligible to participate, individuals were required to have a cancer diagnosis and be undergoing or have undergone treatment, and provide informed consent. Participants were also required to be primary English speakers with at least a basic level of digital experience (assessed via a screening questionnaire). If individuals had participated in any market research in the last 6 months or worked in web design, IT, market research, media or broadcasting, then they were not eligible to participate. Participants were offered a monetary incentive for participation. We aimed to recruit eight participants (which was deemed appropriate to meet the needs of usability testing) aged 40–70 years (with representation in the following age groups: 40–55, 55–65 and 65–70), a balance of males and females, and a range of oncology types and stages of treatment. The intention was that this would increase the likelihood that participants interviewed would provide a range of views and experiences.

Interview methods

Prior to conducting the interviews, the study research team, composed of user experience and eCOA specialists, developed a semi-structured interview guide. This detailed the most important questions to address during the interviews, while allowing participants to elaborate on key issues of specific relevance to their experience.

One-to-one semi-structured interviews based on this predefined interview guide were carried out in February 2019 at a UX research lab in London, UK by a qualitative researcher independent from the solution developer commissioning this research. Interviews were 90 min in duration and were video recorded. During the interview sessions, participants were given a tablet and mobile device containing Signant Health's TrialMax eCOA software solution and asked to interact with the technology as if they were participating in a clinical trial. Sessions were streamed to an observation room where field notes were also taken. No repeat interviews were conducted.

Qualitative data analysis

Video recordings of the interviews were analyzed by an eCOA scientific researcher (FM) who was independent from those who designed or conducted the interviews. Interview recordings underwent manual thematic analysis following the recommended phases: familiarization, generating initial codes, searching for themes, reviewing themes, and defining and naming themes [21,22]. To increase reflexivity, members of the wider scientific team (all possessing research qualifications) contributed to analysis by providing review and sense checking of the derived themes.

Results

Participant characteristics

See Table 1 for participant characteristics.

Eight participants took part in the interviews, one of whom was excluded from analysis as during the interview they reported not yet having started treatment and so the information they provided was not applicable to the research question (their demographic information is not reported here). Thus, the current study analysis included seven individuals (4 females and 3 males) aged 30–68 years with a cancer diagnosis who had undergone or were currently undergoing treatment. Participants had been diagnosed with a range of cancer types and their time since diagnosis ranged from 3 months to 8 years. All participants reported having access to either a smartphone, tablet or both, and all described themselves as having at least an 'intermediate' experience with technology.

Emerging themes

Overall, participants had a positive attitude toward the use of technology to complete assessments and expressed (unprompted) a preference for electronic over paper completion. Furthermore, mobile devices were reported as a helpful tool to stay connected when confined to the home during periods of illness and/or treatment.

Themes that emerged in analyses with implications for device use were: peripheral neuropathy of the hands, tiredness and fatigue, difficulties maintaining focus and concentration, and varying health status based on timepoint in the treatment cycle. See Tables 2 and 3 for supporting quotes.

Peripheral neuropathy of the hands

57% (4/7) of participants reported experiencing peripheral neuropathy of the hands that would impact their interaction with the devices used for recording ePROs. Descriptions of how this manifested included a lack of feeling in the hands, numbness, pins and needles, and stiffness (See Table 2). When treatment is complete, the

Table 1. Participant characteristics.

Characteristics	n (7)
Sex	
– Male	3
– Female	4
Age (years)	
– Mean	54.75
– Range	30–68
Highest education level achieved	
– Left education at 16 years old	2
– Left education at 18 years old	2
– Undergraduate	2
– Postgraduate	1
Employment status	
– Part time	3
– Full time	3
– Retired	1
Oncology diagnosis	
– Breast†	3
– Prostate	2
– Colon/bowel	2
Digital access	
– Smartphone only	4
– Tablet only	0
– Both	3
Digital experience	
– Intermediate	3
– Experienced	4

†Including one metastatic breast cancer.

peripheral neuropathy does not necessarily subside, and dexterity issues can remain long term. One participant reported that they used their smartphone less due to treatment affecting their hands and fingers, and another described not being able to use their hands straightaway in the morning.

However, participants did not see this as a barrier to completing the PRO instruments electronically and viewed using electronic reporting as superior to paper. With one participant saying it is "*110% easier*". Furthermore, one participant reported difficulty writing during treatment and another reported difficulty turning pages due to numbness.

Regarding the implications of peripheral neuropathy in the hands for specific response scale types associated with common PRO instruments, there was no overall consensus on whether participants would prefer a numeric rating scale (NRS) or a sliding scale (visual analogue scale/VAS with a sliding marker). Some participants commented that a scale with larger buttons would be preferable (see Table 3) and that numbness would make it harder to use a sliding scale.

Fatigue and/or concentration issues

86% (6/7) of participants reported fatigue and/or concentration issues that may impact their ability to complete PRO instruments (of note, this would not be unique to electronic implementation). This could last throughout the day, for the duration of the treatment period, and for some time after treatment was complete. Participants did not report issues reading *per se*, but rather that they were tired and fatigued which meant they sometimes lacked the ability to interact mentally with reading materials and it could affect their concentration. One participant reported that they could probably only do a few questions at a time due to the fatigue. However, despite a reduced interaction with technology, another reported that even at the worst time they would still be able to use their phone. Importantly, no readability issues with the information on the devices were reported.

When asked about whether they had experienced 'chemobrain', only 29% said, 'yes' (2/7). Specifically, one participant described suffering short-term memory loss which could mean they would require more frequent reminders of the recall period for questions if there were multiple questions relating to this. They also reported heavily relying on images during chemotherapy instead of only written information. Another said they may not remember training instructions as a result of 'chemobrain'.

Timepoint in treatment cycle

71% (5/7) of interviewees detailed that the timepoint in their treatment cycle would influence their health status and perceived ability to efficiently complete ePROs, with the early stages of treatment cycles reported as when they were feeling their worst (see Table 2). For example, one participant commented that they did not send post/mail during their first round of chemotherapy. Another said that they were fine on the day of chemotherapy but the second day they were weak and then they slept for two days, so they may find it difficult to complete instruments during this time. However, most participants said it is likely that they could find some time during the day when they would be able to complete ePROs.

Table 2. Representative participant quotes for the emerging themes impacting device use.

Emerging themes impacting device use	Supporting quotes
Peripheral neuropathy of the hands	"*I kind of had no feeling in my hands…it was a numbness and a stiffness*" "*When you've just got no energy and you're starting to get the mobility problems…more the numbness, not so much the pain…sometimes you just cannot use your hands…that can last for a day or two…it wouldn't be a problem though…it would still be possible…you're not so bad that there is nothing you can do that whole 24 hour period*" "*I had numbness, I still have numbness…At the beginning, I think the very first session, it did affect me, but I do have numbness now still…but it doesn't affect my holding*" "*The numbness in the fingers makes precise actions very difficult and frustrating.*" "*I'd find all precise interactions very difficult and frustrating, if not impossible*" "*Even if you do have problems, you can still tap*" "*I prefer to use a phone…sometimes you don't want to do all the writing and it's more easy to tap*"
Fatigue and/or concentration issues	"*Fatigue hits you out of nowhere*" "*You are so fatigued*" "*I just felt very tired, exhausted…I don't think it would be quite a case of switching off, but it would certainly be a significant reduction*" "*Since I've had cancer, my reactions and things have slowed down…it slows you down…it's the illness and the treatment*" "*Sometimes your concentration is not too good*" "*You are slower, because you can ask me questions and it might take me a second or two to click in to respond to you…at the first or second session*" "*I did find it difficult to concentrate cause it does make you very tired…it probably reduced my interaction with the phone*" "*The reading and concentration was impacted by the tiredness…it was pretty much throughout [the day]…pretty tired most of the time*" "*Going back to the tiredness…I am mentally more sluggish in the evening, now I don't know if that's age or medication…or just tiredness…thinking of the timing of this, that's why maybe it's better to have the beginning and the end of the day, because some people are night people and they are better at night*"
Timepoint in treatment cycle	"*On the first session I was very ill, you have it every two weeks the chemo, but after a certain period you get to adapt to it…I could still text I just couldn't really hold a conversation*" "*Days 3–6 were my 'dark' days and I did not leave my bed, speak to anyone or even eat any food.*" "*Unless you you've got someone else who you can say 'can you press the button for me?'…sometimes you can be in great pain….no no that wouldn't cause any great problems…it would be alright…the chemo affects everyone differently….I did have a bit of problems with it but that was just in the early days if you had told me to do this everyday it would be like 'oh no', but I think I would have found some way of doing it though, so no, no it wouldn't be a hardship*"

Suggested solutions to improve usability

Encouragingly, despite the side effects of the treatment that could affect their interaction with technology, participants did not feel this would deter or inhibit them from completing PRO instruments on the devices.

"*You can be bad, but not so bad that you can't use nothing at all. . . I would make the effort*"
"*It's not difficult*"
"*You'd make time in the day to do it*"
"*I would still do it*"

However, there were some suggested approaches to improve usability directly related to the identified themes. These included: larger, well-spaced buttons to deal with finger numbness, the ability to pause a survey mid-way and complete at a later point, and presenting the recall period with every question to reduce reliance on memory. Reminders would also be important, and if they had paused in-between questionnaires it would be important to prompt them to return and complete the remaining ones. See Table 3 for supporting quotes.

Participants were also asked about some potential solutions to assist them with completing the ePROs (see Table 3). One such solution was to have caregivers assisting with completion; this was generally not a preferred solution, with one participant not wishing to worry their partner by letting them know how bad they are feeling.

One participant, however, did feel that caregiver assistance could be a beneficial option at times. When asked about the use of a voice assistant, some were open to this idea but most tended not to see this as preferable as it could be annoying. There were mixed views about using a stylus, with most feeling they would not like it, although the option to have it would be appreciated. Finally, some participants mentioned that they would like to have an app on their own phone as opposed to a provisioned device which could just be another thing to remember; they may not take a provisioned device everywhere with them and so at certain times of the day they may miss the time window to complete the ediary (See Table 3).

Table 3. Representative participant quotes in relation to solution optimization and corresponding themes.

Theme	Supporting quotes for solutions to improve usability
Peripheral neuropathy of the hands	"*Why are these so small when there is space to make them bigger?*" (in reference to some of the selection boxes on a scale) "*The bigger the buttons, the better*"
Fatigue and/or concentration issues	"*If you can do something like this on your own time that would be better'* "*It's quite prohibitive actually to have such a short time window… but by actually setting the rules like that, you are actually giving a degree of importance to the survey… that's a positive*" "*Sometimes you can't do everything all at once… sometimes you might feel okay to start doing it and then all of a sudden… urgh… so it would be nice to know you could go back to it and save it, that would help*" "*If you are not in the right frame of mind… that is very important*" (when asked about being able to save questionnaires and go back to them)
Timepoint in treatment cycle	"*Maybe at different times, maybe I woke up in the morning, I didn't want to get out of bed, I was ill, I had a bad night, but in the afternoon at 4 o'clock maybe I felt a bit better*" "*Maybe you can save what you've done and go back into it, that would be good… somebody could be sick*" "*I think there should be something for if you can't fill it in, cause sometimes you don't feel like doing anything that day… chemo is normally in the mornings… sometimes you might not feel like doing anything in the morning, sometimes you want to do something in the evening… I think you should be able to go in when you can… sometimes you cannot do things in set times cause you don't have the energy… the chemo takes it out of you and you don't want to do anything until night time, and sometimes you're all buoyed up before your chemo and you can do it in the morning… your health is up-and-down*" "*Save it and go back to it… if you are not a well person like I wasn't*"
Caregiver assistance	"*No, I'd want it done myself*" "*Yeah that would be a help actually… if you could give it to someone else, if you had to*"
Voice assistant	"*How many mistakes does it make? It makes mistakes… I don't trust it… If you don't read back the text, you've sent something which is nonsense*" "*Even if you have no energy, you still talk, you might be a bit slow… that could be a help… especially if you cannot use anything but you can still talk*"
Stylus	"*I can't stand them… maybe the first or second session, maybe that would have helped me… I wouldn't have kept it though*" "*I might like it… I'd probably opt for a stylus… I'm not strong about it but I would probably if I had an option… to be more accurate*" "*That could help, but if you had problems holding it wouldn't do anything for you but it's good to have, but sometimes you cannot hold anything anyway… even if you are hurting you can still pick up to hold… so yeah it would help… I think it would be easier*"
Use of own device	"*What I would prefer… is to have an app on my own phone*" "*I found it easier to use a laptop*"

Discussion

Participants in the current study reported that the following anticancer treatment related factors would impact their completion of ePRO assessments in a clinical trial: peripheral neuropathy of the hands, experiencing significant tiredness and fatigue which can also impair concentration, and the timepoint in their treatment cycle. The identified themes do not act independently of one another: peripheral neuropathy can be a direct side effect of anticancer treatments, but could also be an indirect symptom resulting from cognitive impairments that affect psychomotor function [9]; fatigue can significantly impact cognition and concentration [13]; and peripheral neuropathy and tiredness and/or fatigue, as well as difficulties with concentration may be more likely to affect individuals at certain times of the day, as well as be influenced by where they are in the treatment cycle. Every patient is different, and a one-size fits all approach is unlikely to be suitable for all. Therefore, identifying potential solutions to optimize implementation of ePROs for as many individuals as possible is desirable, and being flexible in the approach is likely to have the most beneficial effect.

Participants reported that the CIPN they experienced would not be an insurmountable barrier to them completing PRO instruments electronically, and they saw electronic reporting and 'tapping' as superior to writing or turning pages. However, participants commented that larger buttons, the functionality to choose between tapping or sliding an anchor (where individuals would move the anchor up-and-down by maintaining contact with the screen) on a visual analogue scale, and the option of a stylus could assist completion. With knowledge of the high prevalence of

CIPN, the absence of an effective treatment [9,23], and a move toward recording PROs electronically, it is important to take this into consideration when designing electronic solutions.

In line with the literature documenting the commonality of cancer-related fatigue, the participants interviewed reported that their cancer and treatment had led to substantial tiredness, fatigue and/or concentration issues that may impact their ability to complete PRO instruments, and could sometimes lead to a reduced ability to interact mentally with reading materials. It is important to note that this would have an impact on PRO instruments in general and not be specific to electronic implementation, and participants did not report readability issues with the information presented on the devices they were shown.

Some participants reported that they would still be able to use their phone even when they were experiencing their fatigue at its worst, and others said it is likely that they would only be able to complete a few questions at a time. Participants also reported that the impact of their symptoms varied during the treatment cycles, with the beginning of the cycle tending to be the time when they were at their worst, which is in line with other studies that have shown the first few days following treatment to be when symptoms (e.g., fatigue) are at their peak [24]. Encouragingly, participants in our study reported that they would be able to find some time in the day when they could complete the ePROs. However, at such times, they may feel unable to complete PROs on an at-home ediary within a restricted time window and require more flexibility in both the time of day and the time window that they can complete them. Further, such symptoms can occur suddenly, emphasizing the importance of flexibility in the daily schedule of ediary completion.

It is recognized that enabling flexibility in the time window to complete measures and allowing participants to pause and return to a survey may be seen as problematic to trial sponsors, as it could impact the contemporaneous nature of the data which is used to ensure the validity of group comparisons. However, it is important to weigh this up against the higher likelihood of missing data if this is not an option. In their guidance on the evaluation of anticancer medicinal products, the EMA recognize that PRO measures should be administered [4]:

"*when it is feasible to expect high levels of completion by the individual*"

Some participants also described memory deficits and so it would be important that the recall period for questions is reiterated if multiple questions relate to that recall period across screens. Reminders for ediary completion would also be important, especially if the individual was not ready to complete it at the beginning of the time window. Additionally, simple solutions to reset forgotten PINS and access to quick 'how to' guides would be important due to memory issues that can be experienced.

When considering ways to improve ease of participation in clinical trials with at-home electronic data collection, some participants commented that they may like the option to complete ePROs on their own devices (phone, laptop) – known as 'Bring Your Own Device' (BYOD). Participants will be familiar with their own devices and so this option could increase engagement and compliance, potentially reducing missing data [25]. Further, measurement equivalence between electronic formats has been demonstrated [25,26], and so having participants complete the same assessment on different electronic devices would not impact the data being collected. Voice-assisted technology and caregiver completion tended not to be desirable solutions; some participants reported not wanting their 'carer' to know how sick they were really feeling, and so it is important to consider possible issues regarding 'truthfulness' of answers provided in this scenario [27].

Encouragingly, participants reported that the symptoms described would not deter them from participating in clinical trials and they had a positive attitude toward the use of technology to complete assessments. Completing measures electronically was preferable to paper completion and deemed more suitable given some of the impairments that can result from treatment. Nevertheless, our findings do suggest ways that ePRO implementation in oncology clinical trials can be optimized, and given the increasing drive for measuring outcomes that are important to the patient, applying this learning as we design oncology-specific solutions is imperative. Based on our findings we recommend consideration of the implementation best practices listed in Table 4 when using ePROs in this therapeutic population.

It is important to note these recommendations are 'where possible', as not all instrument authors will grant permission for changes to the presentation of scales when migrated from paper to electronic format. However, increasing accessibility is imperative and it is hoped that instrument authors will be aware of the challenges faced by oncology patients (given many of the scales used will be oncology-specific) and willing to facilitate their optimization to ensure they are fit-for-purpose.

Table 4. Recommendations for best practice ePRO implementation in oncology clinical trials.

	Recommendation
a)	Utilize larger, well-spaced buttons.
b)	Enable functionality to choose between tapping or sliding an anchor (where individuals would move the anchor up-and-down by maintaining contact with the screen) on a visual analogue scale
c)	Present recall period along with the full question text and response options to reduce the reliance on participants needing to remember the question recall period from previous screens.
d)	Enable some flexibility in the time window that participants complete the survey/questionnaire.
e)	Allowing 'rest' intervals between questionnaires so that participants can return to remaining questionnaires to complete at a later point (the time window should be adequately detailed in the trial protocol).
f)	Provide a stylus so that participants have this option.
g)	If suitable (taking into account trial design), implement PRO measures on the participant's own device via an app or web-based solution (BYOD).
h)	Automated completion reminders.
i)	Be sensitive that participants may prefer to not complete PROs, rather than asking for caregiver assistance.
j)	Provide simple approaches for participants to reset their own PIN if forgotten, and to access simple quick reference guides.

BYOD: Bring Your Own Device; PRO: Patient-reported outcome.

Along with the increase in implementing PRO instruments electronically in clinical trials, there has also been an increase in trials incorporating measurements from wearable devices and other sensors [28]. These can be used to monitor a range of metrics, including sleep (duration and efficiency), activity, heartrate and blood pressure. The passive, nonobtrusive nature of their measurement offers a particularly attractive option for certain patient populations who are at high risk of treatment-related side effects (such as oncology), providing substantial benefit by reducing participant burden [29]. That is not to say that they would replace PROs, rather wearables could compliment these [30], offering some indication of health status when participants are too unwell to complete PRO instruments, such as tracking impact and recovery profiles associated with each treatment cycle. Such objective data could also help validate and corroborate the data from PRO measures [30]. Furthermore, the data obtained can be very informative about the impact of treatment during the early stages of a treatment cycle, which can be masked by the end-of-cycle site-based assessment when health status may have improved [30]. Studies have shown that in oncology patients step count is associated with likelihood of hospitalization, adverse events and mortality, and a strong predictor of treatment-related toxicity [30]. Studies investigating sleep patterns demonstrated that large discrepancies between self-report and actigraphy-based capture of habitual sleep patterns highlight the value of wearables when investigating such variables in relation to the efficacy and effectiveness of oncology treatments [31].

Adoption of these solutions in clinical trials is sometimes associated with uncertainty by researchers on whether regulators will accept sensor/wearables data in regulatory decision-making and drug labeling claims [28]. However, as PROs are increasingly used in drug approval applications, it is likely that data from wearables and sensors will also grow [28]. Future studies exploring the feasibility and acceptability of these solutions in this patient population would be valuable.

To our knowledge, this is the first study in oncology exploring how the disease and symptoms associated with its treatment can impact device use for recording ePRO data. Participants represented a broad range in age and cancer types and varying times since diagnosis and treatment length. We recognize the limitations of our study, including the small sample size (as is common in qualitative studies); however, participants were consistent in the issues they discussed and reported. Additionally, we acknowledge that the issues raised in these interviews may not reflect the issues experienced by all people with cancer and that these interviews were part of a broader usability study – a context which can encourage participants to be overly critical, and over think potential issues which may otherwise not have been reported. Nevertheless, our findings are in line with previous studies in the literature. Future studies in larger samples will be beneficial to strengthen the findings from this study.

Conclusion

Electronic capture of PROs offer considerable benefit for oncology patients in clinical trials but our findings highlight that peripheral neuropathy of the hands, tiredness, fatigue and/or concentration impairments experienced as a result of the disease and side effects of treatment would have specific implications for their use of technology for completing ePROs. While the identification of these symptoms is not new, the patient perspective on how they would impact their use of technology and strategies for optimizing usability is, and to our knowledge has

not been explored. The insight gained in this research can be harnessed to ensure electronic solutions are optimal and 'fit-for-purpose', and the lessons learned used to inform best practice recommendations for implementation of ePROs in this population, placing patient needs at the forefront. We encourage clinical trial sponsors to take these into consideration when designing and implementing ePRO solutions in oncology studies.

Summary points

- There is an increasing drive in clinical research to measure outcomes that are important to the patient, and electronic implementation of PROs (ePROs) offer considerable benefit for oncology patients in clinical trials. However, symptoms associated with the disease and treatment can impact their use of technologies for recording PROs.
- To our knowledge, this is the first study in oncology exploring how the disease and symptoms associated with its treatment can impact device use for recording ePRO data.
- Participants reported that during the course of the disease they experienced chemotherapy-induced peripheral neuropathy of the hands, tiredness, fatigue and/or concentration impairments, and varying health status based on where they were in the treatment cycle, which would impact their device use.
- Participants had a positive attitude toward the use of technology to complete assessments and completing measures electronically was preferable to paper completion.
- We recommend consideration of the following when implementing ePRO measures in this therapeutic population: utilizing larger, well-spaced buttons; enabling functionality to choose between tapping or sliding an anchor (where individuals would move the anchor up-and-down by maintaining contact with the screen) on a visual analogue scale; presenting the recall period together with the question text and all response options to reduce the reliance on participants needing to remember the question recall period from previous screens; enabling some flexibility in the time window that participants complete the survey; allowing 'rest' intervals between questionnaires so participants can return to remaining questionnaires to complete at a later point; providing a stylus so that participants have this option; if suitable, implementing PRO measures on the participant's own device via an app or web-based solution 'Bring Your Own Device'; automated completion reminders; being sensitive that participants may prefer to not complete PROs, rather than asking for caregiver assistance; and provide simple approaches for participants to reset their own PIN if forgotten, and to access simple quick reference guides.
- In combination with the previous literature demonstrating significant side effects of anticancer treatments, the findings from this research can facilitate the development of patient-centered solutions that account for the unique challenges experienced by oncology patients, and ensure electronic solutions are optimal and 'fit-for-purpose'. The best practice recommendations for implementation of ePROs in this population provide the potential to improve data quality and places patient needs at the forefront.

Author contributions

B Sanderson and B Byrom contributed to the conception and design of the study. FD Mowlem analyzed the data and wrote the manuscript. All authors contributed to interpretation of the findings, reviewed and edited drafts of the manuscript, and approved the final version.

Acknowledgments

The authors extended their thanks to all the participants who gave up their time to take part in these interviews and provided valuable insights.

Financial & competing interests disclosure

When conducting this work, all authors were employed by Signant Health. Otherwise, the authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1. Basch E, Abernethy AP, Mullins CD *et al.* Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. *J. Clin. Oncol.* 30(34), 4249–4255 (2012).
2. Fiero MH, Roydhouse JK, Vallejo J, King-Kallimanis BL, Kluetz PG, Sridhara R. US Food and Drug Administration review of statistical analysis of patient-reported outcomes in lung cancer clinical trials approved between January, 2008, and December, 2017. *Lancet Oncol.* 20(10), e582–e589 (2019).

3. Gnanasakthy A, Barrett A, Evans E, D'Alessio D, Romano C, De M. A review of patient-reported outcomes labeling for oncology drugs approved by the FDA and the EMA (2012–2016). *Value Heal.* 22(2), 203–209 (2019).
4. EMA. EMEA Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. 44(April), 1–18 (2014).
5. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. *Br. Med. J.* 324(7347), 1193–1194 (2002).
6. FDA. Guidance for industry Patient Reported Outcome Measures: Use in medical product development to support labeling claims. 1–39 (2009). www.fda.gov/media/77832/download
7. Giesinger JM, Wintner LM, Zabernigg A *et al.* Assessing quality of life on the day of chemotherapy administration underestimates patients' true symptom burden. *BMC Cancer* 14(1), 1–7 (2014).
8. Kelley MR, Fehrenbacher JC. Challenges and opportunities identifying therapeutic targets for chemotherapy-induced peripheral neuropathy resulting from oxidative DNA damage. *Neural Regen. Res.* 12(1), 72–74 (2017).
9. Zajaczkowską R, Kocot-Kępska M, Leppert W, Wrzosek A, Mika J, Wordliczek J. Mechanisms of chemotherapy-induced peripheral neuropathy. *Int. J. Mol. Sci.* 20(6), 1451 (2019).
10. Brown TJ, Sedhom R, Gupta A. Chemotherapy-induced peripheral neuropathy. *JAMA Oncol.* 5(5), 750 (2019).
11. Ebede CC, Jang Y, Escalante CP. Cancer-related fatigue in cancer surviorship. *Med. Clin. North Am.* 101(6), 1085–1097 (2017).
12. CRUK. Fatigue and cancer drugs (2020). www.cancerresearchuk.org/about-cancer/cancer-in-general/treatment/cancer-drugs/side-effects/fatigue
13. Pendergrass JC, Targum ST, Harrison JE. Cognitive impairment associated with cancer: a brief review. *Innov. Clin. Neurosci.* 15(1-2), 36–44 (2018).
14. Janelsins MC, Kesler SR, Ahles TA, Morrow GR. Prevalence, mechanisms, and management of cancer-related cognitive impairment. *Int. Rev. Psychiatry* 26(1), 102–113 (2014).
15. Gutmann DH. Clearing the fog surrounding chemobrain. *Cell* 176(1-2), 2–4 (2019).
16. Hurria A, Somlo G, Ahles T. Renaming 'chemobrain'. *Cancer Invest.* 25(6), 373–377 (2007).
17. Moore HCF. An overview of chemotherapy-related cognitive dysfunction, or 'chemobrain'. *Oncol. Willist. Park* 28, 797–804 (2014).
18. Ahles TA, Saykin A. Cognitive effects of standard-dose chemotherapy in patients with cancer. *Cancer Invest.* 19(8), 812–820 (2001).
19. Abernethy AP, Herndon JE, Wheeler JL *et al.* Feasibility and acceptability to patients of a longitudinal system for evaluating cancer-related symptoms and quality of life: pilot study of an e/tablet data-collection system in academic oncology. *J. Pain Symptom Manage.* 37(6), 1027–1038 (2009).
20. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad. Med.* 89(9), 1245–1251 (2014).
21. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. *Int. J. Qual. Methods* 16(1), 1–13 (2017).
22. Braun V, Clarke V. Using thematic analysis in psychology. *Qual. Res. Psychol.* 3(2), 77–101 (2006).
23. Staff NP, Grisold A, Grisold W, Windebank AJ. Chemotherapy-induced peripheral neuropathy: a current review. *Ann. Neurol.* 81(6), 772–781 (2017).
24. de Jong N, Kester AD, Schouten HC, Abu-Saad HH, Courtens AM. Course of fatigue between two cycles of adjuvant chemotherapy in breast cancer patients. *Cancer Nurs.* 29(6), 467–477 (2006).
25. Byrom B, Gwaltney C, Slagle A, Gnanasakthy A, Muehlhausen W. Measurement equivalence of patient-reported outcome measures migrated to electronic formats: a review of evidence and recommendations for clinical trials and bring your own device. *Ther. Innov. Regul. Sci.* 53(4), 426–430 (2018).
26. Byrom B, Doll H, Muehlhausen W *et al.* Measurement equivalence of patient-reported outcome measure response scale types collected using bring your own device compared to paper and a provisioned device: results of a randomized equivalence trial. *Value Heal.* 21(5), 581–589 (2017).
27. Abma JC, Chandra A, Mosher WD, Peterson LS, Piccinino LJ. Fertility, family planning, and women's health: new data from the 1995 National Survey of family growth. *Vital Heal. Stat.* 19, 1–114 (1997).
28. Cox SM, Lane A, Volchenboum SL. Use of wearable, mobile, and sensor technology in cancer clinical trials. *JCO Clin. Cancer Informatics* (2), 1–11 (2018).
29. Purswani JM, Ohri N, Champ C. Tracking steps in oncology: the time is now. *Cancer Manag. Res.* 10, 2439–2447 (2018).
30. Gresham G, Hendifar AE, Spiegel B *et al.* Wearable activity monitors to assess performance status and predict clinical outcomes in advanced cancer patients. *NPJ Digit. Med.* 1(1), 1–8 (2018).
31. Moore CM, Schmiefe SJ, Matthews EE. Survivors: discrepancy in selected sleep parameters. *Behav. Sleep Med.* 13(6), 472–490 (2015).

Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Journal of Comparative Effectiveness Research

Guidance for a causal comparative effectiveness analysis emulating a target trial based on big real world evidence: when to start statin treatment

Felicitas Kuehne[1], Beate Jahn[1], Annette Conrads-Frank[1], Marvin Bundo[1], Marjan Arvandi[1], Florian Endel[2], Niki Popper[2,3,4], Gottfried Endel[5], Christoph Urach[4], Michael Gyimesi[6], Eleanor J Murray[7,8], Goodarz Danaei[8,9], Thomas A Gaziano[10,11], Ankur Pandya[10] & Uwe Siebert*[,1,10,12,13]

[1]Department of Public Health, Health Services Research & Health Technology Assessment, Institute of Public Health, Medical Decision Making & Health Technology Assessment, UMIT – University for Health Sciences, Medical Informatics & Technology, Hall iT, Austria

[2]DEXHELPP, Vienna, Austria

[3]TU Wien, Research Unit of Information and Software Engineering, Austria

[4]dwh GmbH, Simulation Services & Technical Solutions, Austria

[5]Department EWG, Main Association of Austrian Social Security Institutions, Vienna, Austria

[6]Austrian Public Health Institute, Austria

[7]Department of Epidemiology, Boston University School of Public Health, Boston, MA 02118, USA

[8]Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA 02115, USA

[9]Department of Global Health & Population, Harvard TH Chan School of Public Health, Boston, MA 02115, USA

[10]Department of Health Policy & Management, Center for Health Decision Science, Harvard TH Chan School of Public Health, Boston, MA 02115, USA

[11]Division of Cardiovascular Medicine, Brigham & Women's Hospital, Boston, MA 02115, USA

[12]Cardiovascular Research Program, Institute for Technology Assessment & Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115, USA

[13]Division of Health Technology Assessment & Bioinformatics, ONCOTYROL – Center for Personalized Cancer Medicine, Innsbruck, Austria

*Author for correspondence: Tel.: +43 50 8648 3930; uwe.siebert@umit.at

Aim: The aim of this project is to describe a causal (counterfactual) approach for analyzing when to start statin treatment to prevent cardiovascular disease using real-world evidence. **Methods:** We use directed acyclic graphs to operationalize and visualize the causal research question considering selection bias, potential time-independent and time-dependent confounding. We provide a study protocol following the 'target trial' approach and describe the data structure needed for the causal assessment. **Conclusion:** The study protocol can be applied to real-world data, in general. However, the structure and quality of the database play an essential role for the validity of the results, and database-specific potential for bias needs to be explicitly considered.

First draft submitted: 1 October 2018; Accepted for publication: 5 July 2019; Published online: 12 September 2019

Keywords: big data • cardiovascular disease • causal inference • inverse probability of censoring weighting (IPCW) • observational study • real-world evidence • statin • study design • target trial • time-dependent confounding

Background

Statin treatment in the prevention of cardiovascular events

There is an ongoing debate on the optimal use of statin treatment in the prevention of cardiovascular events. In 2016, the European Society of Cardiology (ESC) published new guidelines for the prevention of cardiovascular disease [1]. These guidelines recommend lipid-lowering treatment such as statins for excessive-risk individuals with overt cardiovascular disease, diabetes, chronic kidney disease, familial hypercholesterolemia, or very high levels of individual risk factors. Outside this excessive-risk group, the ESC Systematic Coronary Risk Evaluation

Future Medicine

(SCORE) [2] is recommended for total disease risk assessment [1]. This country-adapted SCORE system intends to assess a patient's 10-year risk of undergoing a fatal cardiovascular event. The risk estimation is based on age, sex, total cholesterol (with an optional use of high-density lipoprotein cholesterol), smoking status and blood pressure, and it should be repeated every 5 years. Patients are divided into four risk categories, low (<1% SCORE), moderate (1–5% SCORE), high (5–10% SCORE) and very high (>10% SCORE). The target for low-density lipoprotein cholesterol (LDL-C) levels and the recommended treatment and treatment dose depend on the risk category and the LDL-C level [2]. Immediate statin treatment is recommended for a risk score ≥10 and an LDL-C level ≥70 mg/dl or a risk score ≥5 and <10, and an LDL-C level ≥100 mg/dl. However, for all risk scores ≥5, drug treatment could be considered [1,2].

The American College of Cardiology and the American Heart Association (ACC/AHA) released new recommendations in November 2013 [3], in which the threshold indication for statin therapy was lowered for primary prevention. Instead of LDL-C thresholds, these recommendations focus on total atherosclerotic cardiovascular disease (CVD) risk, which is defined by new Pooled Cohort Equations [3]. Statin treatment is recommended when the atherosclerotic CVD risk equals or exceeds 7.5%. Pandya estimated the cost–effectiveness of various risk thresholds in the US American setting and concluded that more lenient thresholds of 4% or higher would be optimal using a threshold of $100,000 per quality adjusted life year (QALY) gained [4].

This raises the question whether there is evidence that a policy oriented purely on the risk score and that is using a more lenient threshold would also be beneficial in the European setting. Vancheri and others could not demonstrate an association between an increased statin utilization and coronary mortality or its rate of change in European countries between 2000 and 2012 [5]. However, this study did not have access to data on adherence, indication for statin treatment and industry campaigns on doctor's decision making, which could have confounded the results. Adherence to statin treatment for instance, is known to be negatively associated with mortality as well as with the intensity of statin regimen [6,7].

Challenges in real-world data: causal inference

Real-world evidence (RWE) is a widely propagated source to gain information outside the artificial setting of clinical trials [8–10]. However, these data bear their own difficulties, including confounding, missing or nonreliable data, no clear treatment assignment, imperfect compliance, dynamic treatment regimes, switching and multiple-line treatments. Several approaches exist that may help conducting studies that allow drawing causal conclusions from observational studies and RWE data [11]. Visual, structural and statistical techniques exist to understand and estimate causal relations using RWE [11–42].

Directed acyclic graphs are graphical models that help in visualizing causal relations. In addition, causal-directed acyclic graphs help to understand the nature of confounding, to assess identifiability, and to guide the selection of the appropriate statistical analytic approach [18,22,32–37,39,41,42].

A prominent structural approach is the so-called 'target trial' approach described by Hernan, Robins, Cain and others [11,13–15,23,25] combined with statistical methods controlling for time-dependent confounding. The target trial approach suggests designing any observational data analysis as if one would design a randomized clinical trial. This approach is consistent with counterfactual theory in the analysis of causality, and additionally, it is structurally close to the randomized trial approach, and therefore, facilitates the communication to medical researchers familiar with clinical trial design. In addition to comparison of treatments, this approach allows researchers to answer the question of when to start a specific treatment (e.g., when to start statin) [14].

Causal statistical methods include the (parametric) g-formula, marginal structural models with inverse probability weighting and structural nested models with g-estimation. These are complex and validated statistical methods that can be applied when time-dependent confounding is present. These techniques are all based on the counterfactual approach and differ regarding their assumptions [12,13,16,17,19–21,24,26–32,38,40]. In the counterfactual approach, one estimates the outcome assuming the individual had received no treatment.

A combination of these visual, structural and statistical causal methods allows estimating causal relations using RWE and is particularly useful to determine risk thresholds for initiation of preventive medication such as statin use.

Aim

The overall goal of this project is to demonstrate how causal inference methods can be applied to RWE to obtain unbiased estimates. We focus on statin treatment to prevent CVD events in individuals with and without a history

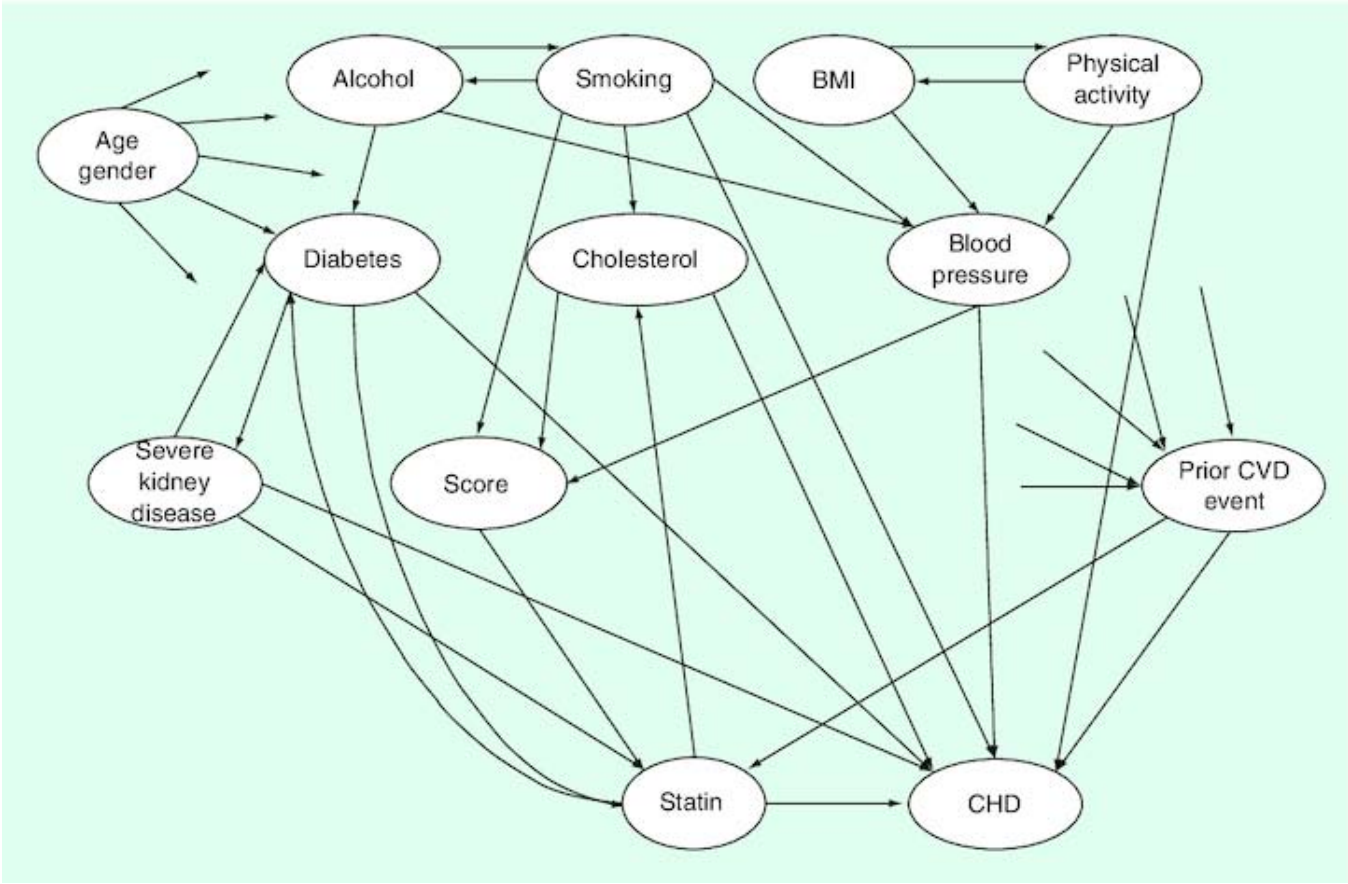


Figure 1. Causal graph. Arrows between the variables represent potential causal effects. Age and sex may have an influence on all of the parameters, and prior cardiovascular disease events may also be influenced by all parameters other than 'statin' and 'coronary heart disease'. For space reasons, those arrows are only indicated. The absence of an arrow represents the assumption of the absence of any causal effects.
BMI: Body mass index; CHD: Coronary heart disease; CVD: Cardiovascular disease.

of a cardiovascular event. In the absence of a clinical trial evaluating the optimal start of statin treatment, we would like to guide the use of available real-world data and causal inference methods. We aim to provide a study protocol to answer the question of when to start statin treatment to optimally prevent CVD events. It is not the intention of this paper to present the results of the data analysis.

Methods

For the development of the study protocol, we operationalized the research question, developed a causal graph to identify potential confounders and followed the target trial approach [11,14,15,23,25].

Causal graph

Real-world data analyses are particularly prone to confounding and selection bias. Unlike in a randomized clinical trial, the intervention – in our case the (time of starting) statin treatment – is not administered at random but based on individual clinical information. When the same factors that influence the decision whether or not treatment should be provided, also influence the outcome, in our case CVD events, controlling for confounding is essential for obtaining valid causal effect measures. To provide the basis for the selection of variables needed to be controlled for in a causal analysis, we generated a causal graph including variables that are directly or indirectly associated with the intervention statin treatment and the outcome CVD (Figure 1). The assumptions on the presence or absence of associations were based on published literature [43–47] and expert opinions. We searched the literature for CVD-risk evaluations and risk prediction models as well as for clinical treatment guidelines. We then asked the clinical experts which factors they base their decision on when starting or not starting statin treatment. We included those factors

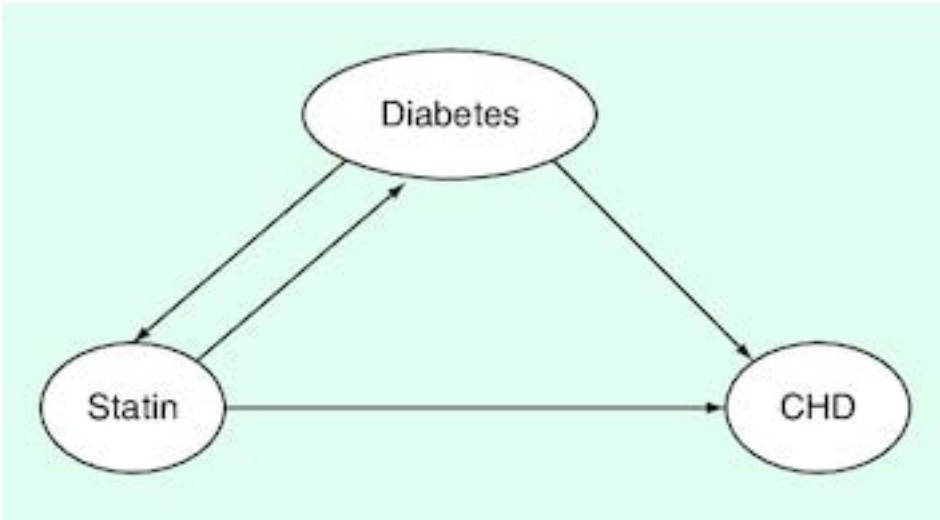


Figure 2. Time-dependent confounding.
Diabetes is a confounder and intermediate step, a so-called time-dependent confounder.
CHD: Coronary heart disease.

into the causal graph irrespective of their availability in the datasets. Some factors such as birth date and sex may not change after the treatment decision has been made and can be included as baseline variables. However, other variables, such as any disease-specific and treatment variables may change over time and should be included in the analyses as time-varying variables. The assumed causal direction is indicated by the arrows. Some variables are connected by links pointing in both directions. Those cases represent relationships with feedback loops over time.

The causal graph (Figure 1) shows that confounding can be eliminated by blocking the backdoor paths from statin treatment to cardiovascular events, that is, controlling for age, sex and the ESC SCORE, or when the ESC score is not available, all the variables being included in the ESC SCORE as well as diabetes and severe kidney disease. Similarly, as the variables included in the risk score, the variables severe kidney disease and diabetes may change over time, and therefore, should be included in the model as baselines variables and time-varying variables.

The ESC SCORE, and newly diagnosed severe kidney disease and newly diagnosed diabetes are time-dependent confounders [24,32]. On one side, they are influencing the probability of physicians prescribing the start of statin treatment, and on the other side, they are themselves affected by statin treatment (Figure 2). This creates a feedback loop between time-dependent treatment and time-dependent confounder, and hence, causal methods such as g-methods [32] need to be applied to draw valid (unbiased) causal conclusions.

Target trial protocol

In the absence of a randomized clinical trial comparing multiple strategies for when to start statin treatment, we based our protocol on the target trial approach, which emulates a trial using real-world observational data: when designing a target trial, one formulates the study protocol as if one was designing a randomized controlled trial. This approach assures that the timely sequence and analytic strategy is appropriate and biases such as immortal-time bias are avoided [25]. The target trial emulates (mimics) a randomized controlled trial, and therefore consists of the same components as a randomized controlled trial. These components include eligibility criteria, treatment strategies, assignment procedures, follow-up period, outcomes, causal contrasts of interest (i.e., the estimands), and analysis plan and start with a well-formulated research question [11].

For our statins study, each component of the protocol was carefully considered and defined according to the operationalized research question. As result, we describe each component of the target trial study protocol.

Operationalization of the research question

To assess the optimal (i.e., most effective) timing of starting statin treatment, several comparative strategies with different treatment starting points must be defined. Currently, the statin treatment strategy is based on the ESC

SCORE, which assesses the individual 10-year risk of undergoing a fatal cardiovascular event based on age, sex, total cholesterol, smoking status and blood pressure, and the LDL-C level [1]. Our intention is to define the treatment strategies by describing the starting point of statin treatment entirely based on ESC SCORE thresholds.

As in a randomized clinical trial, the individuals in the compared strategies should be as similar as possible. Individuals are therefore included in each strategy at the time of study inclusion. Just the defined treatment starting point differs. Figure 3 shows that while being in the study, the risk may increase as time passes by, and the individual may cross the specific risk threshold of a certain given strategy. At that time point, treatment should start.

The operationalized research question is shown in the following population, intervention, comparison and outcome (PICO) structure:

Population: Men and women aged 40–75 years of the general Austrian population.

Interventions:

- Immediate statin treatment when risk score equals or exceeds 1%.
- Immediate statin treatment when risk score equals or exceeds 2%.
- Immediate statin treatment when risk score equals or exceeds 3%.
- Immediate statin treatment when risk score equals or exceeds 5%.

Comparator: No statin treatment.

Outcomes: Primary: CVD defined as composite end point major adverse cardiovascular event (MACE), that is, myocardial infarction (MI) or stroke, or cardiovascular death; secondary: overall survival.

This research question reflects the study population, treatment strategies and outcomes as defined below.

Figure 3. Treatment strategies. Time does not reflect the risk score increase proportionally. The colors reflect the treatment as defined for each strategy in the protocol. Light gray indicates that the individual has a risk score below the given risk threshold and should not take statin treatment; dark gray indicates that the individual risk levels crossed the risk threshold of the corresponding strategy and the individual should be treated with statins. The shaded gray indicates the grace period, which is 3 months after crossing the risk threshold, in which an individual may start the stating treatment. A given subject not following the protocol will be censored at the time of protocol violation.
tx: Treatment.

Study population & eligibility criteria

Following the target trial concept, only those individuals are included into the analyses who could be randomized to one of the four treatment strategies or the comparator. Hence, we plan to include the general population between the age of 40 and 75 years who are not classified yet as excessive risk, that is, no overt CVD, diabetes, chronic kidney disease, familial hypercholesterolemia, very high levels of individual risk factors or an acute cardiovascular event. Individuals are included into the study when they first show to have crossed the 10-year risk of undergoing a fatal cardiovascular event threshold of 1%. Note that inclusion in the study does not mean immediate statin treatment as treatment depends on the strategy-specific threshold.

Our exclusion criteria are age <40 years or >75 years, diagnosis of stroke ≤1 month prior of the start of the emulated trial (ICD10: I60–64), diagnosis of MI ≤1 month prior of the start of the emulated trial (ICD10:

I21–25), diabetes (ICD10: E10–14), chronic kidney disease (ICD10: N18) and familial hypercholesterolemia (ICD10: Z82)

Individuals with and without a history of prior cardiovascular events are included in the study. The analyses will be conducted in subgroups: without prior cardiovascular events (primary prevention); and with prior cardiovascular events (secondary prevention).

Treatment strategies

Five treatment strategies (i.e., four statin strategies and the comparator no statin treatment) will be compared, differing in whether and when to start statin treatment as shown above.

All individuals enter the study at a similar state of risk, that is, when they first show to have crossed the 10-year risk of undergoing a fatal cardiovascular event threshold of 1%. This implies that individuals in the treatment strategies who start statin treatment at a high-risk threshold wait and will be monitored until they cross the risk threshold corresponding to the particular treatment strategy they are assigned to (Figure 3).

The statin treatment used in the observational study is defined by prescription of any available statin filed in the database. The dose of statin treatment is not considered for treatment definition. By applying the prescription of statins as indicator of statin use, it is assumed that individuals take the medication as it was prescribed.

Assignment procedure

Subjects will be assigned to a treatment arm at the beginning of the study. Rather than randomizing subjects to a single treatment arm, so called 'replicates' ('copies' or 'clones') of each subject in the dataset will be created. One replicate of each subject is then analytically being allocated to each of the treatment strategies, and subjects not following the assigned treatment protocol will be censored at the time of protocol violation. For example, for a subject A starting treatment with a risk score of 3%, the subject's data are copied five-times and each copy is then allocated to a different treatment strategy at the beginning of the study. In the 1% threshold strategy, the replicate of subject A would then be censored right at the beginning of the study, as the protocol is to start treatment when crossing the 1% risk threshold and subject A does not start until the risk exceeds 3%. In contrast, in the 2% threshold strategy, the data of the replicate of the same subject A are censored at the time she/he has a 2% risk of cardiovascular event. As treatment is not started, it is violating the 2% threshold protocol. In the 3% threshold strategy, the data of the replicate of subject A are not censored, as treatment is started at a 3% risk, which is following the protocol. In the 5% threshold strategy and the no-treatment strategy, the data of the replicate of subject A are censored at the time of treatment start, as this is the time when the observation indicates assigned treatment violation. As censoring is informative, adjustment for baseline and time-dependent confounders of the censoring–outcome relation must be performed to derive valid causal treatment effects and is described in the analysis plan.

Individuals assigned to one of the four immediate treatment arms receive statin treatment after crossing the corresponding risk threshold (Figure 3). We allow for a 'grace period' of 3 months during which the treatment may be started after crossing the respective threshold, as individuals may not start treatment immediately due to nonmedical reasons, such as vacation, family holidays, travel and other personal reasons. Individuals not starting statin treatment within this grace period will be censored at the end of the grace period (i.e., 3 months after crossing the risk threshold). Individuals who had started treatment according to the protocol and stop it for more than 3 months are also planned to be censored 3 months after the last statin intake. The impact of the length of the grace period on the results should be tested in sensitivity analyses.

The individuals in the no-treatment arm must not receive statin treatment at any time. Therefore, individuals in this arm who start treatment at any time are being censored at the time of first statin use.

Follow-up periods

The study baseline is at the time of eligibility, which is the time point when an individual's ESC-SCORE exceeds 1%. The study's follow-up period starts at baseline and ends at the end of the 15-year follow-up time, occurrence of the outcome, loss of follow-up or treatment violation, whichever occurs first.

Outcomes

Our primary end point (outcome) is defined as any major cardiovascular event (MACE), including MI, stroke or death due to CVD during the time of follow-up. Our secondary end point is death from any cause during the time

of follow-up. As frequency measures to be compared between the treatment arms, we use the 15-year incidences (risks) of the outcomes.

Causal contrast(s) of interest

Although a clinical trial may be analyzed to estimate the intention-to-treat effect, this is usually often not possible when using existing real-world data. In contrast, when using the target trial approach, all subjects are replicated and each subject is assigned to each treatment arm. Subsequently the causal per protocol effect of the respective sustained treatment strategy is estimated, that is, the effect of the assigned treatment strategy if all individuals in the trial had sustained statin treatment until end of study as indicated in the protocol. As censoring due to treatment violation or censoring due to loss to follow-up maybe informative, methods must be applied that account for informative censoring (see below).

Analysis plan

Descriptive analyses

Descriptive statistics will summarize the data. The descriptive statistics should start with a flowchart clearly indicating the number of individuals included and excluded by the various reasons. The study population will then be described within each treatment arm and subgroup. As all subjects are assigned to each treatment arm, not all individuals assigned to the treatment arms will be described but those that follow the protocol, and those that are censored. These statistics will describe age, sex, the presence of each risk factor, comorbidities and survival for adherent and censored individuals within each treatment group and compare between groups.

Causal analyses

The causal analysis will be performed for the entire population as well as the subgroups with and without prior cardiovascular events, with inverse probability weighting with marginal structural models [13,15,17,19,21,23,26,29–32,40]. According to the causal graph, the analysis will be adjusted for age, sex, prior cardiovascular events, blood pressure and cholesterol as baseline variables and cholesterol, new onset of severe kidney disease and diabetes as time-dependent confounders [17,28,29,31,32].

Time intervals will be created to properly adjust for time-dependent confounding. The length of the interval depends on the available data. The time-varying variables must be defined within each interval. Hence, intervals need to be short enough to avoid bias through time-dependent confounding.

One replicate of each subject is assigned to each of the compared treatment arms, in order to emulate perfect randomization [11,25]. Within each treatment arm, each subject will be censored at the time when this subject stops following (i.e., violates) her/his assigned treatment protocol. In other words, for each subject, only information is used that is compatible with the assigned treatment at each time point. In order to allow for a causal interpretation of the analytic results, the fact that censoring due to protocol violation may be informative needs to be considered. Therefore, adjustment for informative censoring by performing an analysis with inverse probability of censoring weighting will be performed. Individuals will be weighted using the predicted probability of not being censored due to protocol violation in a given time interval in each assigned treatment arm. The inverse of that probability is the assigned weight for that person interval. A low probability of not being censored, therefore, results in a high weight. As very high weights (e.g., values >100) may generate unstable results, the distribution of estimated weights will be assessed.

The intervals are set at 3 months, as data collection mechanisms are often implemented per quarter due to quarterly reimbursement. When the data indicate that those intervals are too long because variables are measured several times within one interval or too short as no changes occur between intervals, the length of the intervals may be adjusted and explored in sensitivity analyses.

Each subject per interval will have an assigned weight, which is estimated using a model predicting the start of statin treatment, the so called 'treatment selection model'. This probability will then be used to indicate the probability of being censored in a given treatment strategy and turned into a weight. This weight will then be used in the so called 'outcome model', which relates the treatment arm to the outcome. In the outcome model, the study population hazard ratios for MACE for each treatment strategy will be approximated from a marginal Cox proportional hazards model with the odds ratio from a pooled logistic regression comparing MACE between assigned treatment arms within intervals [19]. Robust variance estimators will be used to adjust for dependence between different person time (intervals) of subjects; and the replicates generated from each particular subject. In

addition, we will compare the 15-year proportions (cumulative incidences) of CVD events and deaths between all arms.

Missing data

In many datasets, variables that are based on ICD10 codes may have no indicator whether they contain missing variables or not. In the case of no data entry, one could assume that no diagnosis is present or that the information is missing. This introduces unmeasured confounding and may be systematic when this information is not entered for a specific patient group or reason related to the outcome.

Depending on the database, diagnosis data may not be well monitored and therefore not valid. In those cases, other variables such as medication use, age and sex could be used to predict and impute diagnoses [48]. In cases where no diagnosis data (e.g., ICD10 code) indicate the absence of disease but the prediction based on medication indicate a probability of disease >80%, it will be assumed that the disease is present. This assumption needs to be tested in sensitivity analyses. To reflect the uncertainty associated with the estimation of the missing data, multiple imputations will be performed [49].

Many variables of our study have repeated measurements. These variables might not have values for each time point. The method of the last value carried forward will be applied assuming that the physician deciding on the treatment bases the decision on the last available information.

Data sources

To follow the study protocol, several variables need to be available at baseline (e.g., age, sex, and prior diagnoses of CVD, diabetes, chronic kidney disease, and familial hypercholesterolemia). Variables that require repeated measures include diagnoses, laboratory values, behaviors and measurements of blood pressure. Depending on the data source, not all of these data might be available. Continuous information on laboratory values, behaviors and clinical measurements are often lacking. This information could be indirectly estimated or approximated using data on diagnoses or on filled prescriptions.

If not all data are available from one source, the necessary information may be obtained by linking several databases containing administrative reimbursement data, specialized registry data, clinical information and operative data.

Discussion

As statin treatment is a widely accepted and reimbursed treatment for the prevention of CVD events, it is unlikely that an expensive trial will be designed to find the best starting point of statin treatment. However, real-world data on statin treatment and outcomes are available from routine clinical practice. Conclusively, these data could be used to find the optimal treatment strategy. We describe a causal approach to use RWE to answer the causal question of when starting statin treatment is optimal from the patient perspective. Although using real-world data to estimate the comparative effectiveness of statin treatment avoids the artificial setting of a clinical trial and reflects reality better than clinical trials, these analyses are only valid if complete information of baseline and time-dependent confounders is available. We described the data necessary for a causal analysis of observational data. This information should be used when making decisions on the observational database that is analyzed.

In medical practice, data are often collected for various reasons, including assessing the long-term safety and effectiveness of different regimens in routine clinical care, as well as issues with adherence. However, payment or reimbursement often is the primary reason for data collection and data on diagnoses, procedures, prescriptions, and visits reflect the billing procedures but not necessarily the clinical process and reality. Using these data for clinical analyses comprises some challenges. First, clinical and reimbursement data may diverge slightly in primary versus secondary diagnoses. Second, big real-world data are not monitored closely and time intervals are irregular. Missing data are, therefore, difficult to identify.

The target trial concept is relatively innovative and only few conceptual publications exist describing this approach [11,13–15]. The target trial concept differs from disease area to disease area. Before applying the target trial concept, we used the causal graph (Figure 1) to identify variables that must be included into the analysis as baseline or time-varying covariates and based the target trial protocol on those findings. This preparatory step is the basis for controlling for confounding and selection bias. To our knowledge, no target trial protocol exists yet for the comparison of different when to start statin treatment strategies using real-world databases. The evaluation of real-world data becomes especially essential as machine learning (ML) and artificial intelligence (AI) are becoming

increasingly important in medicine. However, ML and AI can only be successful when the data are reliable and biases are avoided by applying the appropriate analytical methods based on explicit causal assumptions.

The suggested approach requires comprehensive data, including ambulatory, laboratory, behavioral information and measurements of blood pressure. High-quality data are essential for drawing valid and reliable conclusions. European countries provide only limited access to such data. However, we know that data are collected in almost all parts of life including medicine. On the other hand, generic availability of statins has resulted in a change in statin prescribing behavior. Hence, the consequences of different statin prescription patters are reflected in such data. If and when these data will be made available, the described protocol can be applied. As analyses following this type of protocol will provide good insight on the effectiveness of different treatment strategies at a much lower cost than a clinical trial, it should be debated whether necessary data could be linked and made accessible without violating patients' rights.

This proposed protocol for the design and analysis of an emulated target trial using observational RWE has several limitations. When defining the study population, we excluded individuals with a cardiovascular event within 1 month prior of start of emulated trial. This washout phase might be too short. Also, prior statin treatment and prior cardiovascular events might not always be identifiable. Thus, a few individuals might be classified as having a risk score of 1% when it is actually higher.

Data that are included in the ESC SCORE are crucial for the proposed analyses. If these data are not reliable, the analyses may be biased. We might not only have misclassification when allocating the treatment strategies, we would also not correctly adjust for confounding. As the causal graph (Figure 1) shows, the ESC SCORE is influencing the treatment and is itself influenced by the treatment. Hence, the ESC SCORE is a time-dependent confounder, where the influence of the risk score on the treatment needs to be controlled for. The physician is calculating the risk in the practice based on the direct measurements of blood pressure cholesterol level and other patient-specific information. We might not have the ESC SCORE value that the physician sees and have to estimate it using the available underlying data. When this leads to misclassification, the link between the score and the treatment initiation is not correctly analyzed.

We intended to compare five treatment strategies (Figure 3). Theoretically, many more treatment strategies could be compared even if they seem unethical. This proposed target trial is applied to observational data. Hence, the actual treatment of an individual is not influenced by the protocol. However, whenever analyzed strategies diverge from current guidelines, it might be questionable whether a sufficient number of cases follow the protocol of those treatment strategies. Data must be sufficient to predict protocol violation. For cases where only few individuals have the same covariates that predict treatment, the weights would become very large; this is one of the limitations of inverse probability of censoring weighting. To avoid large weights, one could truncate the weights. However, large databases reflecting all treatment arms and covariate strata are always preferable and yield more reliable results.

The primary study outcome chosen in our protocol is MACE, which includes cardiovascular death. However, the databases do not always differentiate between all-cause mortality and sudden cardiovascular death. When a subject died due to a cardiovascular event before reaching the hospital, death may only be considered as all-cause mortality.

The protocol demands an immediate start of statin intake after crossing the risk threshold corresponding to a given treatment strategy. In the real world, individuals may want to delay the start of the treatment due to personal nonmedical reasons. We account for that by allowing for a grace time of 3 months after crossing the risk threshold indicating treatment start and test the impact of the length of the grace period with sensitivity analyses. During this grace time period, individuals are supposed to start the indicated treatment. Only individuals that have not started treatment by the end of the grace-time period are perceived as violating the protocol.

Often, study protocols include independently adjudicated outcome validation to avoid potential differential information bias. However, we believe that MACE is a relatively well-defined outcome leaving no room for too much interpretation. Furthermore, at the time of data collection, no connection between the different statin treatment strategies and the outcome was made. Hence, the misclassification would potentially be nondifferential. We also included overall survival as secondary end point in our analysis, as this outcome is less prone to information bias. However, the disadvantage is that due to limited number of deaths in population studies, the power for differentiating between the slightly different strategies is very limited.

Although we design the observational study as if it was a clinical trial, the trial should be seen as a pragmatic trial, where the treatment strategies reflect the real-world conditions. Such a trial can be analyzed according to

the causal per-protocol principle for sustained treatment strategies, which avoids some biases when compliance is incomplete [50].

Currently, access to comprehensive data including data from the outpatient setting is limited in Europe. The FFG COMET project, Decision Support for Health Policy and Planning (DEXHELPP), aims to develop new methods, models and technologies to support analysis, planning and control of healthcare systems [51–53]. One part of the project is to extend causal inference methods in order to estimate causal relationships from routine data, especially for the parameterization of decision-analytic models [54]. Within the project DEXHELPP, large Austrian databases are linked to RWE on a per-project basis as well as a service for data owners as stable foundation for various research purposes. Differing sources including closed, nonpublic databases with highly sensitive data as well as publicly available, open data are utilized. 'GAP-DRG' (General Approach for Patient oriented Outpatient-based diagnosis-related group [DRG]) [55,56], a database built up and managed on behalf of the Main Association of Austrian Social Security Institutions, is one of the outstanding assets of the project. It comprises of administrative claims data from many sections of the healthcare system as hospital discharges, ambulatory outpatient contacts with general practitioners (GPs) and specialists as well as filled in prescriptions and leaves of absence. DEXHELPP and preceding projects linked these databases and added additional information, meta-data and documentation.

After designing the study protocol, the next step is to perform the study for the Austrian population. Once the data and their structure accessible through the DEXHELPP consortium meet the necessary data quality for the proposed protocol, this protocol will be applied to estimate the most efficient start of statin treatment for the Austrian setting.

Conclusion

We successfully developed a study protocol for estimating the comparative effectiveness of different starting points for statin treatment to prevent cardiovascular events. The study protocol follows the target trial approach to allow drawing causal conclusions from real-world observational databases. It can be applied to clinical as well as reimbursement databases or the combination of both. The quality of the databases plays an essential role for the reliability of the results. The proposed study protocol has some limitations that need to be acknowledged when applying it to any data and interpreting the results. We conclude that the proposed protocol can be applied to different databases to estimate the comparative effectiveness of different starting points for statin treatment to avoid cardiovascular events. However, the database this protocol will be applied to needs to be carefully selected, and database-specific potential for bias must be carefully considered.

Future perspective

Over the last decades, computers and electronic tools arrived in basically all parts of our lives, including medicine. The storage of health-related data is increasing, and potentials of linking different databases are growing, analytical techniques are emerging, and ML and AI in medicine become a widely debated topic [57]. We believe that in 5–10 years, artificial intelligence will be more and more accepted in several parts of medicine. Especially, risk prediction and causal inference will facilitate accurate personalized medicine. However, the potentials of future performance of artificial intelligence depend on current research in the field of causal inference from real-world data.

Financial & competing interests disclosure

The research project DEXHELPP (Decision Support for Health Policy and Planning: Methods, Models and Technologies based on Existing Health Care Data, grant no. 843550) is in the frame of COMET-Competence Centers for Excellent Technologies. DEXHELPP is supported by Federal Ministry of Transport, Innovation and Technology (BMVIT), Federal Ministry of Science, Research and Economics (BMWFW) and the state Vienna. The COMET program is transacted by the Austrian Research Promotion Agency (FFG). In parts, this work has also been financially supported through Erasmus Mundus Western Balkans (ERAWEB), a project funded by the European Commission. The funding body did not have any influence on the design of the study, and collection, analysis and interpretation of data, and in manuscript writing. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Summary points

- It is questionable whether a more lenient threshold for the use of statin treatment than currently proposed in European guidelines leads to better clinical outcome.
- Common challenges occurring with real-world evidence are confounding, missing/misclassified data, no clear treatment strategy assignment, dynamic treatment regimens and treatment switching.
- The aim of this project is to describe a causal (counterfactual) approach for analyzing real-world observational data to estimate the optimal risk threshold for the start of statin treatment.
- Visual, structural and statistical techniques existing to estimate causal relations using real-world evidence are applied to generate a study protocol.
- We operationalize the research question, describe the data structure needed for the causal assessment and address potential biases that might occur.
- Using causal graphs, we identified age, sex, prior diagnosis of cardiovascular disease, diabetes, chronic kidney disease and familial hypercholesterolemia as baseline (time-independent) confounders. We found new onset of diabetes or chronic kidney disease, cholesterol and blood pressure as time-dependent confounders.
- Crucial parts of the protocol are the treatment allocation at the beginning of the trial, a per-protocol analysis with correct censoring and a statistical analyses accounting for informative censoring such as inverse probability of censoring weighting.
- The proposed protocol can be applied to different databases to estimate the comparative effectiveness of different starting points for statin treatment to avoid cardiovascular events. The database-specific potential for bias needs to be estimated and considered.

References

Papers of special note have been highlighted as: • of interest

1. Catapano AL, Graham I, De Backer G *et al.* 2016 ESC/EAS guidelines for the management of dyslipidaemias. *Eur. Heart J.* 37(39), 2999–3058 (2016).
2. Perk J, De Backer G, Gohlke H *et al.* European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). *Eur. Heart J.* 33(13), 1635–1701 (2012).
3. Andrus B, Lacaille D. 2013 ACC/AHA guideline on the assessment of cardiovascular risk. *J. Am. Coll. Cardiol.* 63(25 Pt A), 2886 (2014).
4. Pandya A, Sy S, Cho S, Weinstein MC, Gaziano TA. Cost–effectiveness of 10-year risk thresholds for initiation of statin therapy for primary prevention of cardiovascular disease. *JAMA* 314(2), 142–150 (2015).
5. Vancheri F, Backlund L, Strender LE, Godman B, Wettermark B. Time trends in statin utilisation and coronary mortality in western European countries. *BMJ Open* 6(3), e010500 (2016).
6. Deshpande S, Quek RG, Forbes CA *et al.* A systematic review to assess adherence and persistence with statins. *Curr. Med. Res. Opin.* 33(4), 769–778 (2017).
7. Rodriguez F, Maron DJ, Knowles JW, Virani SS, Lin S, Heidenreich PA. Association of statin adherence with mortality in patients with atherosclerotic cardiovascular disease. *JAMA Cardiol.* doi:10.1001/jamacardio.2018.4936 (2019) (Epub ahead of print).
8. Berger ML, Sox H, Willke RJ *et al.* Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Value Health* 20(8), 1003–1008 (2017).
9. Greenfield S. Making real-world evidence more useful for decision making. *Value Health* 20(8), 1023–1024 (2017).
10. Wang SV, Schneeweiss S, Berger ML *et al.* Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. *Value Health* 20(8), 1009–1022 (2017).
11. Hernan MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* 183(8), 758–764 (2016).

• **Introduces the 'target trial' framework. It discusses each component of the target trial.**

12. Bellamy SL, Lin JY, Ten Have TR. An introduction to causal modeling in clinical trials. *Clin. Trials* 4(1), 58–73 (2007).
13. Cain LE, Caniglia EC, Phillips A *et al.* Efavirenz versus boosted atazanavir-containing regimens and immunologic, virologic, and clinical outcomes: a prospective study of HIV-positive individuals. *Medicine (Baltimore)* 95(41), e5133 (2016).
14. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *Int. J. Biostat.* 6(2), Article 18 (2010).
15. Cain LE, Saag MS, Petersen M *et al.* Using observational data to emulate a randomized trial of dynamic treatment-switching strategies: an application to antiretroviral therapy. *Int. J. Epidemiol.* 45(6), 2038–2049 (2015).
16. Cole SR, Hernan MA. Fallibility in estimating direct effects. *Int. J. Epidemiol.* 31(1), 163–165 (2002).

17. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JA. Methods for dealing with time-dependent confounding. *Stat. Med.* 32(9), 1584–1618 (2013).
18. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 10(1), 37–48 (1999).

- **Nicely lays out the principles of directed acyclic graphs. Shows why they are important and discusses unnecessary and harmful adjustment.**

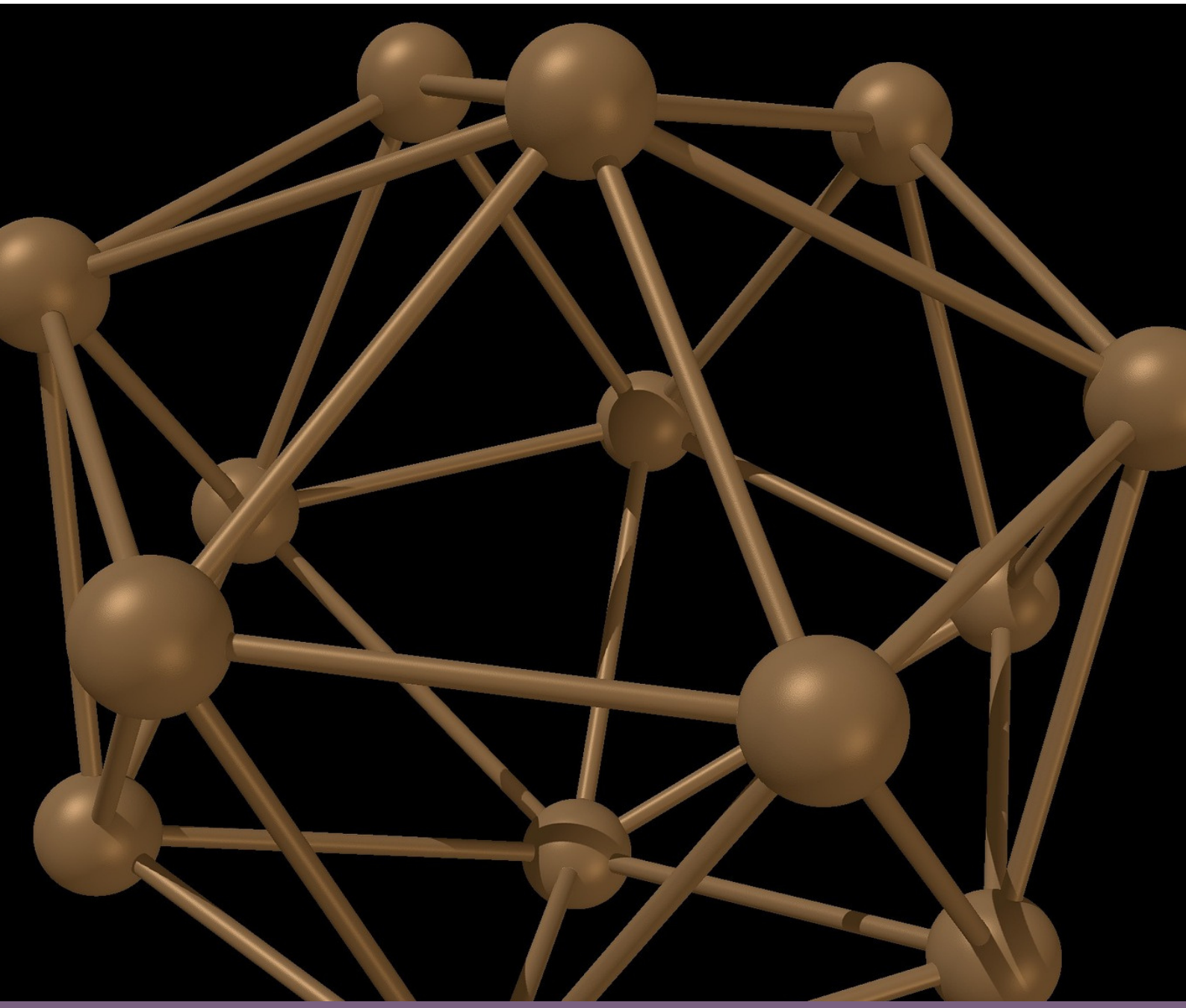
19. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 11(5), 561–570 (2000).
20. Hernán MA, Cole SR, Margolick J, Cohen M, Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol. Drug Saf.* 14(7), 477–491 (2005).
21. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin. Trials* 9(1), 48–55 (2012).
22. Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am. J. Epidemiol.* 155(2), 176–184 (2002).
23. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J. Epidemiol. Community Health* 60(7), 578–586 (2006).
24. Hernan MA, Robins JM. *Causal Inference.* Chapman & Hall/CRC, FL, USA (forthcoming) (2019).
25. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J. Clin. Epidemiol.* 2016(79), 70–75 (2016).

- **Focuses on importance of the correct timing of starting the trial, eligibility and treatment strategy assignment.**

26. Latimer NR, Abrams KR, Lambert PC, Morden JP, Crowther MJ. Assessing methods for dealing with treatment switching in clinical trials: a follow-up simulation study. *Stat. Methods Med. Res.* 27(3), 765–784 (2018).
27. Morden JP, Lambert PC, Latimer N, Abrams KR, Wailoo AJ. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. *BMC Med. Res. Methodol.* 11, 4 (2011).
28. Robins JM. Marginal structural models. In: *Proceedings of the Section on Bayesian Statistical Science*. American Statistical Association, VA, USA, 1–10 (1997).
29. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Springer, NY, USA, 95–133 (2000).
30. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 56(3), 779–788 (2000).
31. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11(5), 550–560 (2000).
32. Robins JM, Hernán MA, Siebert U. Effects of multiple interventions. Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors. 1, 2191–2230 (2004).
33. VanderWeele TJ, Hernán MA, Robins JM. Causal directed acyclic graphs and the direction of unmeasured confounding bias. *Epidemiology* 19(5), 720–728 (2008).
34. VanderWeele TJ, Robins JM. Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *Am. J. Epidemiol.* 166(9), 1096–1104 (2007).
35. VanderWeele TJ, Robins JM. Four types of effect modification: a classification based on directed acyclic graphs. *Epidemiology* 18(5), 561–568 (2007).
36. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. *Biometrics* 67(4), 1406–1413 (2011).
37. VanderWeele TJ, Tan Z. Directed acyclic graphs with edge-specific bounds. *Biometrika* 99(1), 115–126 (2012).
38. Westreich D, Cole SR, Young JG *et al.* The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. *Stat. Med.* 31(18), 2000–2009 (2012).
39. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am. J. Epidemiol.* 177(4), 292–298 (2013).
40. White IR. Methods for handling treatment switching: rank-preserving structural nested failure time models, inverseprobability-of-censoring weighting, and marginal structural models. Workshop: Methods for Adjusting for Treatment Switches in Late-Stage Cancer Trials (2012). www.methodologyhubs.mrc.ac.uk/workshop_summaries/treatment_switches.aspx
41. Williamson EJ, Aitken Z, Lawrie J, Dharmage SC, Burgess JA, Forbes AB. Introduction to causal diagrams for confounder selection. *Respirology* 19(3), 303–311 (2014).
42. Yuan Y, Shen X, Pan W. Maximum likelihood estimation over directed acyclic gaussian graphs. *Stat. Anal. Data Min.* 5(6), doi:10.1002/sam.11168 (2012).
43. D'Agostino RB Sr, Vasan RS, Pencina MJ *et al.* General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 117(6), 743–753 (2008).

44. Fox CS. Cardiovascular disease risk factors, Type 2 diabetes mellitus, and the Framingham Heart Study. *Trends Cardiovasc. Med.* 20(3), 90–95 (2010).
45. Payne RA. Cardiovascular risk. *Br. J. Clin. Pharmacol.* 74(3), 396–410 (2012).
46. Siontis GC, Tzoulaki I, Siontis KC, Ioannidis JP. Comparisons of established risk prediction models for cardiovascular disease: systematic review. *BMJ* 344, e3318 (2012).
47. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 97(18), 1837–1847 (1998).
48. Filzmoser P, Eisl A, Endel F. ATC – ICD: determination of the reliability for predicting the ICD code from the ATC code. Hauptverband der Österreichischen Sozialversicherungsträger, EBM Bericht (2009). www.academia.edu/17416378/ATC_ICD_Determination_of_the_reliability_for_predicting_the_ICD_code_from_the_ATC_code
49. Kang H. The prevention and handling of the missing data. *Korean J. Anesthesiol.* 64(5), 402–406 (2013).
50. Hernan MA, Robins JM. Per-protocol analyses of pragmatic trials. *N. Engl. J. Med.* 377(14), 1391–1398 (2017).
51. DEXHELPP: Decision support for health policy and planning. Comet K-Project DEXHELPP (2018). www.dexhelpp.at/
52. Popper N, Endel F, Mayer R, Bicher M, Glock B. Planning future health: developing big data and system modelling pipelines for health system research. *Sim. Notes Europe* 27(4), 203–208 (2017).
53. Zauner G, Popper N, Breitenecker F. State of the art research in Austria: DEXHELPP – decision support for health policy and planning: methods, models and technologies based on existing healthcare data. *Value Health* 17(7), A452 (2014).
54. Murray EJ, Robins JM, Seage GR 3rd *et al.* Using observational data to calibrate simulation models. *Med. Decis. Making* 38(2), 212–224 (2018).
55. Endel F, Duftschmid G. Secondary use of claims data from the Austrian health insurance system with i2b2: a pilot study. *Stud. Health Technol. Inform.* 223, 245–252 (2016).
56. Endel F, Endel G, Pfeffer N. PRM34 routine data in HTA: record linkage in Austria's GAP-DRG database. *Value Health* 15(7), A466 (2012).
57. Peek N, Combi C, Marin R, Bellazzi R. Thirty years of artificial intelligence in medicine (AIME) conferences: a review of research themes. *Artif. Intell. Med.* 65(1), 61–73 (2015).

The Evidence Base



Contact us

Editorial Department

Editor
Darcy Hodge
d.hodge@future-science-group.com

Business Development and Support

Publisher
Freya Leask
f.leask@future-science-group.com